

NASA needs 'the vision thing'

Planetary scientists and astronomers may fare reasonably well under the US space agency's new budget-conscious chief. But in the long term, can NASA provide the inspiration to excite future generations about these disciplines?

Last week, 41 years to the day after Yuri Gagarin made humanity's first voyage into space, NASA's new administrator Sean O'Keefe went to Syracuse University in upstate New York to deliver his first major policy address. Flanked by two congressmen who chair committees that monitor NASA's activities, O'Keefe outlined a vision that was most notable, well, for its deliberate lack of vision. Don't expect any giant leaps on O'Keefe's watch.

This will be a disappointment to all those who've been tugging on NASA's sleeve ever since the last Moon landing, asking when the agency will head off on its next big adventure. At Syracuse, and at practically every other public forum since taking office in January, NASA's chief has subtly sent the message: we're not sending astronauts to Mars any time soon, so stop asking when.

In part, O'Keefe's hands are tied by fiscal reality. He has been hired by a US administration that wants to bring NASA's budget — in particular that of the International Space Station (ISS) — under control. The current White House admires the pragmatism of corporate managers, and doesn't wax sentimental about exploring space 'because it's there'. So the new administrator talks about building capability and investing in technologies such as nuclear propulsion that will enable somebody else to undertake grand voyages at some point in the indeterminate future.

O'Keefe says that technology and science, not exploration, will be his watchwords. For researchers in space-based astronomy and

planetary science, he will hopefully prove a friendly administrator, steering his agency "where the science dictates that we go", as he put it last week.

But taking a longer view, science could lose out if O'Keefe turns his back entirely on what the first President George Bush once called "the vision thing". Human exploration has a unique romantic appeal. How many of the NASA's current principal investigators were first turned towards a career in research by the excitement of the Apollo programme? And how many young and inquiring minds will be lost to space science if our TV screens are never again lit up with images of astronauts exploring an alien world?

There is still the ISS, of course. But it never set the world of science alight, and as it gets progressively whittled down, it looks less like a viable lab and more like a vehicle for channelling taxpayers' money to aerospace contractors. There's little inspiration to be gained from watching NASA's finest sitting in a tin can.

We may not yet be ready to send astronauts to Mars. But ideas for other, more practical destinations are beginning to bubble up even as the human spaceflight programme faces its worst identity crisis in years. For example, astronauts could be sent on geological field trips to near-Earth asteroids. Such journeys would in many ways be easier than landing on the Moon, and would push the boundaries of human exploration for the first time in 30 years, while returning useful scientific information. ■

Women don't want to be 'one of the boys'

At the top of Japan's scientific establishment, women are faced with a past they thought they had escaped.

At the annual meeting of the Molecular Biology Society of Japan last December, there was a special symposium featuring a panel made up entirely of prominent Japanese women scientists. Held to highlight the achievements of women in research, the symposium provided evidence that Japanese science is making some headway in the quest for gender equality.

Thanks to the courage of prominent women scientists in exposing examples of discrimination (see *Nature* **410**, 404–406; 2001), Japan is facing up to its poor record in promoting women to senior positions. University departments and academic society committees now know that it looks progressive to have a woman among their members. And at increasing numbers of scientific meetings, childcare facilities are provided to make it easier for researcher–mothers to participate.

But how deep is this new-found commitment to women's rights? Not very, argue many of the women who have been let into Japan's upper scientific echelons. They still face accusations that they were promoted just because they were women — and, indeed, many suspect that they received their position as a token gesture. One of the panellists who participated in the December special symposium said she would never do it again. "I am a researcher, not a woman researcher," she told *Nature*.

When it comes to a real understanding of sexual discrimination, or of sexual harassment for that matter, a sizeable part of Japan's

academic establishment — mostly the older portion — just doesn't seem to get it.

This has led to some awkward situations. Official gatherings of senior scientists often have the atmosphere of a men's club. On one level, perhaps, it is encouraging that senior women are now invited to such gatherings. But there is little satisfaction to be gained from attending an event that causes you to squirm with embarrassment. Some among the new class of prominent women researchers, for instance, have been given the 'honour' of participating in outings in which scientists are entertained by geisha.

Geisha are traditional purveyors of song, dance, jokes and various other arts. But there are also sexual overtones to their work, which usually involves entertaining successful men. Leaving aside the issue of whether scientific organizations should be spending the large sums needed to invite a geisha on other, more practical concerns, the organizers of such events should be more sensitive to the attitude of participants.

"It was clearly sexual. It made me feel uncomfortable," laments one researcher of her encounter with a geisha. "But I was polite to her, and she was polite to me."

Unfortunately, Japanese etiquette also required her to be polite to whoever organized the outing. But an insult is an insult, no matter how well-dressed and sweet-sounding. ■





Head start
Fetal transplants
offer relief for
Parkinson's sufferers
p666



Dust storm
Is pollution from
Asia reaching the
shores of the US?
p668



Hot house
Tension in the
Philippines hampers
zoologists
p669



Not guilty
Weapons scientist
walks free from
South African court
p670

Frogs put in the gender blender by America's favourite herbicide

Rex Dalton, Berkeley

Low levels of the most widely used herbicide in the United States have been found to disrupt the sexual development of frogs. The finding will heighten concerns about the persistence of endocrine-disrupting chemicals in the environment — and throws the spotlight on another potential factor involved in the global decline of amphibian populations.

In the United States, about 27,000 tonnes of atrazine are applied each year to fields used to grow crops such as maize, sorghum and sugar cane. It is also used in 80 other nations, making it one of the world's most important herbicides. But concerns about atrazine's potential ability to disrupt sex hormones, and the presence of residues in drinking water, have led it to be banned in Germany, France, Italy, Sweden, Norway and Switzerland.

Tyrone Hayes and his colleagues at the University of California, Berkeley, exposed tadpoles of the African clawed frog (*Xenopus laevis*) to water containing different levels of atrazine throughout their larval development. Levels as low as 0.1 parts per billion (p.p.b.) caused males to develop ovaries in addition to testes. And at concentrations above 1 p.p.b., male frogs' larynges — used to call for mates — failed to develop normally. Hayes reports in a paper published on 16 April (T. B. Hayes *et al. Proc. Natl Acad. Sci. USA* **99**, 5476–5480; 2002).

In addition, when the researchers put adult males in water containing 25 p.p.b. of atrazine, their testosterone levels plummeted. And in preliminary, unpublished research, Hayes has found similar reproductive anomalies in wild leopard frogs (*Rana pipiens*) at six locations in the midwestern United States with high levels of atrazine.

"It is obviously affecting frogs," says Hayes, who suspects that atrazine boosts production of the enzyme aromatase, which catalyses the conversion of testosterone into oestrogen. "We need to ask: what are the environmental costs of using atrazine?" he says.

Although previous studies have suggested that atrazine is an endocrine disrupter, these have mostly used much higher doses. Hayes's study is causing alarm because he has



Do environmental levels of atrazine cause sexual defects in leopard frogs in the US midwest?

observed effects at concentrations that reflect those found in the environment. "Hayes's work is some of the best to show how a contaminant affects amphibian reproduction," says Andrew Blaustein, who studies amphibians at Oregon State University in Corvallis.

"This is not a worst-case scenario where animals are exposed to mega-doses," says Louis Guillette, a zoologist at the University of Florida in Gainesville. "These are concentrations we know wild amphibians are exposed to." Indeed, atrazine is routinely found at

Scripps to trawl sea-microbe data

Jonathan Knight, San Francisco

The genomic secrets of the ocean's microbes are set to be revealed by the world's first centre devoted to marine bioinformatics. The Scripps Institution of Oceanography in La Jolla, California, plans to establish the facility within the next few months.

The centre's main focus will be microbial genomics, says Ron Burton, director of marine biology at Scripps. Most of the marine genomes sequenced so far have been microorganisms — yet their biochemistry and ecology remains poorly understood, despite their crucial role in the global cycling of carbon and nitrogen (see *Nature* **415**, 572–574; 2002).

The Scripps centre will not itself sequence marine genomes, Burton says. That work will continue at places such as the Joint Genome Institute in Walnut Creek, California, where a number of marine microbes have so far been sequenced. Rather,

the new facility's emphasis will be on teasing information about ocean life from the sequence databases.

For example, Brian Palenik's laboratory at Scripps has already started to apply bioinformatics techniques to learn how microbes relate to their ocean environment. Traditionally, biologists have probed this question by studying microorganisms in laboratory flasks. But genomics can help to reveal the factors that contribute to a microbe's survival. Palenik has discovered several overlooked metabolic pathways in cyanobacteria, photosynthetic microbes that are responsible for 80% of the carbon fixed from the atmosphere in the open ocean.

The new division will eventually consist of about 10 faculty members, and will cost upwards of \$20 million to establish. A director has yet to be appointed, but the leading candidate is said to be a well-known figure in the bioinformatics field.

► 50 p.p.b. in water throughout the United States, experts say, and can rise to several parts per million in agricultural run-off.

Amphibians are highly sensitive to pollutants and are regarded as 'sentinel' species that can provide the first indications of damage to an ecosystem. Hayes's results come as the US Environmental Protection Agency (EPA) is re-evaluating the risks posed by atrazine—following a 1999 ruling on a lawsuit from the Natural Resources Defense Council and other environmental groups, requiring the agency to accelerate its efforts to reassess the environmental and health risks posed by a range of potential endocrine disruptors.

The EPA currently allows levels of atrazine to reach 3 p.p.b. in US drinking water. An EPA panel will hold a public meeting this week in Arlington, Virginia, on suggested residue levels of atrazine in food products.

Hayes has provided the EPA with a report on his unpublished field-work on frogs from Colorado to Wisconsin, but EPA officials declined to comment on this, or on Hayes's newly published paper. The agency aims to issue final reports on the risks posed by atrazine by August.

Atrazine is manufactured by Syngenta of Basel, Switzerland. Alan Hosmer, Syngenta's manager of ecological sciences in Greensboro, North Carolina, says that the firm is aware of some of Hayes's results. But the company failed to provide a detailed statement before *Nature* went to press. For three years, Hayes worked on contract research at Berkeley for Syngenta. But he severed relations with the firm last year, independently pursuing the studies that led to this week's paper.

Aside from their implications for the use of atrazine in the United States and elsewhere, Hayes's results add another element to the debate over why amphibian populations are declining across the world. Many factors are thought to be involved, including pathogens, pollution, competition with invading alien species, and habitat loss.

James Hanken, a herpetologist at Harvard University who chairs the international Declining Amphibian Populations Task Force, describes Hayes's results as "compelling". But he notes that some of the most severe declines are in Central America, where atrazine is not thought to be widely used. Hanken argues that further studies are needed to determine what role atrazine and related chemicals are playing in amphibian declines across the globe. ■

► www.open.ac.uk/daptf



Tyrone Hayes has fears for frogs.

Parkinson's patients show positive response to implants

Erika Check, Washington

Fetal cells transplanted into the brains of patients with Parkinson's disease can survive and mature for up to eight years after the surgery, helping to alleviate the debilitating tremors seen in the disease. These positive results, from a trial that last year attracted controversy because of the side-effects experienced by a small number of patients, should help to lift the cloud that has since hung over cellular therapies for the condition.

Parkinson's is caused by the death of cells that secrete the neurotransmitter dopamine. When Curt Freed of the University of Colorado and his colleagues reported on the trial last year, their results confirmed that transplanted cells from the brains of aborted fetuses can help to relieve the symptoms of the disease (see *Nature* **410**, 401; 2001). But five of the 20 patients treated experienced jerky, uncontrolled movements called dyskinesias.

This week, at a meeting of the American Academy of Neurology in Denver, Colorado, Freed reported that these patients have since responded to treatment aimed at addressing this unwanted side-effect—either by giving them drugs to remove excess dopamine or by implanting a device called a 'deep-brain stimulator', which delivers timed electric shocks to the brain region that causes the dyskinesias. Indeed, one of the dyskinetic patients was working in the World Trade Center at the time of the 11 September attacks last year, and was able to make his escape by walking down 33 flights of stairs.

The new results are from a total of 32 patients. The trial was expanded by relaxing its double-blind design and allowing some patients who had initially been in the control group to have the transplants. The degree of improvement varied between patients, but seemed to be independent of age—the team's earlier data had suggested that only those younger than 60 benefited. Patients who did best were those who had previously responded most strongly to the drug L-DOPA, which is converted to dopamine in the brain.

Autopsies on three patients who died up to eight years after surgery showed that the transplanted fetal cells were making increasing amounts of a pigment called neuromelanin. Neurons make neuromelanin as they age. "This says the dopamine neurons we've transplanted appear to be developing as though they were in normal human brain environments," says Freed.

This result is particularly encouraging, says Ole Isacson, a Parkinson's expert at



Fetal cells transplanted into the brains of patients with Parkinson's can ease the disease's symptoms.

Harvard Medical School. "Maturation is the key," he says. "These grafts need to grow and mature to become effective."

But Isacson and other researchers say that cellular therapies will need further refinement if they are to deliver better results than conventional drug therapy. They note that Freed's team transplanted cells into a brain structure called the putamen, which contains the cells that respond to dopamine. But Ivar Mendez, a neurosurgeon at Dalhousie University in Halifax, Nova Scotia, has begun a pilot study in which five patients have received fetal-cell transplants into both the putamen and the substantia nigra—the structure that contains the dopamine-secreting cells lost in Parkinson's.

Mendez also treats fetal cells with a biochemical called glial-cell-line-derived neurotrophic factor before transplanting them. A brain-imaging study indicates that these treated cells survive better in their new environment. Mendez's patients also seem to improve faster than those in Freed's trial.

Other groups, meanwhile, are working on techniques to grow dopamine-producing neurons from cultures of human embryonic or adult neural stem cells, which would remove the need to obtain material from freshly aborted fetuses. ■

Physicists blast US missile defence review

Geoff Brumfiel, Washington

A Pentagon advisory group is considering nuclear-tipped interceptors as a possible alternative to conventional missile defence systems.

But the move has worried scientists studying missile defence technologies. They fear that nuclear interceptors, if ever used, would wipe out vital telecommunications, and that developing them would damage nonproliferation efforts.

Current anti-ballistic-missile systems rely on 'hit-to-kill' strategies, in which interceptors collide with their targets. Hit-to-kill devices destroy nuclear warheads without detonating them—but intercepting a fast-moving missile reliably is technically very challenging, and the devices can be fooled by decoy warheads. Nuclear-tipped warheads could potentially avoid these problems by creating an explosion big enough to knock out any nearby missiles.

Researchers first examined the possibility of using nuclear interceptors in the 1960s and 1970s, but the plan was scrapped, in part because of a lack of public support. "It's been off the charts for 30 years. It deserves re-examination," says William Schneider, who chairs the Defense Science Board, the government advisory group that is conducting the new study. The group plans to review several systems, including nuclear interceptors.

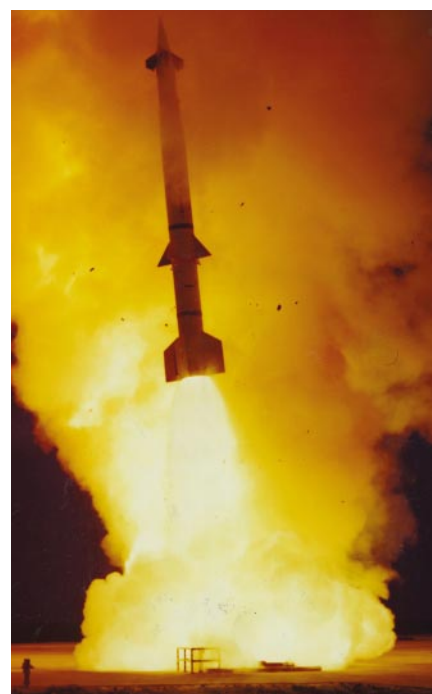
Although the blast produced by the inter-

ceptors is likely to be high enough in space to avoid any fallout on the ground, it would release an electromagnetic pulse powerful enough to knock out communications satellites and cripple electronics on the surface. "This is sort of a desperate, brute-force solution," says Henry Kelly, president of the Federation of American Scientists. "In a country as dependent as we are on advanced communications and electronics, this would have a devastating effect."

Kelly also notes that new nuclear warheads could require experiments that would end the current moratorium on nuclear testing. The National Nuclear Security Administration, which maintains US nuclear weapons, earlier this year announced plans to design new warheads (see *Nature* **415**, 945; 2002). That programme avoids testing, but Kelly believes that more complex nuclear-tipped interceptors would eventually have to be tested.

Department of Defense officials stress that the current review is extremely preliminary. "We do not have any programme for nuclear-tipped interceptors," says Pentagon spokeswoman Cheryl Irwin.

But even the perception that the United States is considering nuclear interceptors could be harmful, according to Steven Weinberg, a physicist at the University of Texas at Austin who studies missile defence technology. "There has been a kind of taboo against



BETTMANN/CORBIS

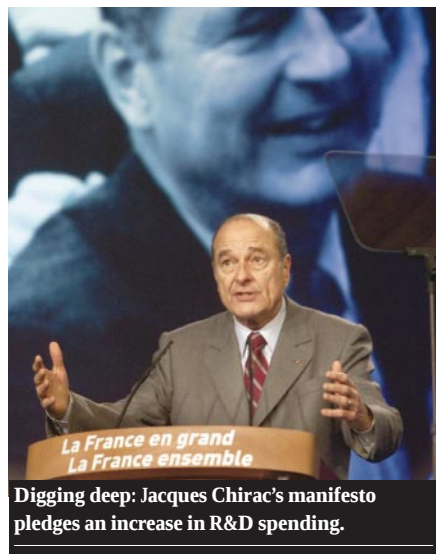
Old idea: plans for nuclear interceptors, such as this Spartan missile, were scrapped in the 1970s.

the use of nuclear weapons — even for peace-time uses. It would be the worst kind of folly for the United States even to be perceived as breaking that taboo," he says. ■

Chirac's scientific overtures greeted by scepticism

Sally Goodman, Paris

France's right-wing president Jacques Chirac has promised to boost innovation and increase public spending on research if he is re-elected.



Digging deep: Jacques Chirac's manifesto pledges an increase in R&D spending.

In a manifesto released to *Nature*, but not yet made public, Chirac pledges to convert the research ministry into a ministry for research and innovation. He wants to boost the mobility of researchers between the public and private sectors and plans tax breaks to encourage corporate R&D investment.

Chirac also argues that public and private spending on R&D should increase from the current 2.1% of French gross domestic product to 3% by the end of the decade. "The commitment to research must be historic," says the manifesto.

Industrial groups welcome Chirac's focus on innovation. But some researchers will be disappointed that the plan does little to address the problems of French universities — which house more than 80% of the laboratories run by public research agencies. Obstacles to mobility and autonomy for young researchers are a particular concern.

Chirac's rival for the presidency, socialist prime minister Lionel Jospin, has pledged to bring universities and research back under one ministry if he wins the election

on 5 May (see *Nature* **416**, 253; 2002).

Many researchers remember their experiences under the last right-wing government, when budgets were cut. "Even if Jospin's track record in terms of research is mixed, that doesn't erase the memory of the 1993–97 period," says Edouard Brézin, a physicist at the École Normale Supérieure in Paris and a former president of the CNRS, France's main agency for basic research.

Gérard Tobelem, a haematologist at the University of Paris VII and national secretary for research in Chirac's RPR party, concedes that research wasn't a priority for the last right-wing government. But he argues that the new plan has "a total awareness of what is now at stake for French science".

In France, the prime minister has ultimate responsibility for science policy. But if parliamentary elections in June go the same way as the presidential election, the new president will be free to appoint a prime minister of his own political persuasion. Chirac and Jospin are currently running neck-and-neck in the opinion polls. ■

Clinton bill calls for revamp of chronic-disease tracking

Virginia Gewin, Washington

Moves are under way to transform the patchy and underfunded system with which US public-health authorities track chronic diseases and their causes.

Plans to upgrade the system were discussed last week at a meeting held at the Institute of Medicine in Washington DC. Delegates have high hopes for the safe passage of the Nationwide Health Tracking Act, which was introduced to Congress last month by Senators Hillary Clinton (Demo-

crat, New York) and Harry Reid (Democrat, Nevada). It advocates spending \$270 million on a new nationwide environmental-health tracking system to integrate local, state and federal public-health schemes, and is likely to attract bipartisan support. US environmental-health activities are currently split between dozens of agencies.

Chronic diseases such as cancer and asthma are leading causes of death in the United States. "Rising rates of chronic disease require us to act now," Clinton told the meeting.

The need to upgrade the monitoring network was raised in 2000 in a report by the Pew Environmental Health Commission, a public-health think-tank based at John Hopkins University in Baltimore, Maryland. Monitoring schemes received \$30 million for the current fiscal year, but experts say that a successful nationwide system would need the larger sum proposed in the Clinton-Reid bill.

The bill proposes establishing five biomedical monitoring laboratories and five environmental-health centres of excellence. The network could also tap into expertise at the Environmental Protection Agency and the US Geological Survey. The latter, for example, already uses remote-sensing satellites to track dust storms that transport heavy metals, pathogens and pesticides.

♦ <http://pewenvirohealth.jhsph.edu>



Hillary Clinton, seen here with a leukaemia patient, is pushing for better disease monitoring.

Money-spinning journal ruse foiled

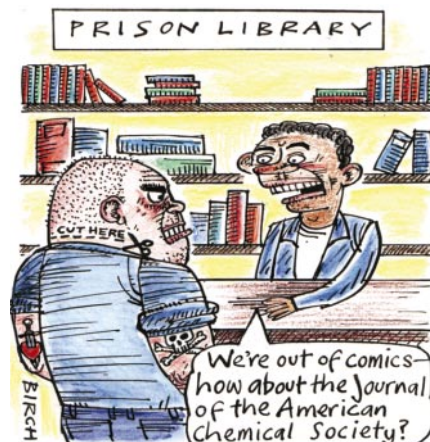
Geoff Brumfiel, Washington

Four scientific societies have rumbled an international scam that is reported to have cost them more than a million dollars.

The operation was run by Eastwood Books, a Los Angeles-based subscription service that renewed subscriptions for university libraries. Manager Jung Shin is alleged to have found a profitable but illicit angle to this mundane service by using fake names and scientific affiliations to purchase individual journal subscriptions. These were then resold to libraries in Asia at institutional rates, which range from two to ten times the individual price.

The four societies — the American Institute of Physics, the American Physical Society, the Institute of Electrical and Electronics Engineers, and the American Chemical Society — announced last week that they had received \$250,000 in an out-of-court settlement after threatening to sue Eastwood Books.

Although this figure is less than their esti-



mated losses, the societies are pleased with the outcome. "The main thing was to stop the practice," says Marc Brodsky, executive director of the American Institute of Physics. Shin declined to comment on the matter, but told *Nature* that she has now left the subscriptions business and resumed her education.



Something in the air? China's pollutants may be exacerbating US health problems.

California observatory sweeps the skies for springtime Asian dust

Rex Dalton, Boulder

As the dust clouds that blow across the Pacific Ocean every spring begin their annual journey, a new observatory is gearing up to assess the impacts of the pollution the clouds bring with them.

The National Oceanic and Atmospheric Administration (NOAA) facility at Trinidad Head, an isolated coastal area some 400 kilometres north of San Francisco, should help to determine whether pollutants from Asia are exacerbating health and environmental problems in the United States.

NOAA has similar observatories in Antarctica, Hawaii, Alaska and American Samoa in the Pacific, but only Trinidad Head is in the correct location to monitor the dust arriving at the west coast of the United States. "We really don't understand how dust and chemicals move and mix globally," says atmospheric scientist Russ Schnell, who directs NOAA's observatory operations from its centre in Boulder, Colorado. "The winds are pushing the pollutants of a billion people out of China over the Pacific."

Dust from China is carried across the Pacific by the easterly winds that blow between March and May, bringing chemical pollutants such as ozone. The dust is generated in China by drought, overgrazing of livestock and poor cultivation practices. China is addressing the dust problem, including a massive effort to plant trees as windbreaks.

Two dozen staff are due to start work this month at Trinidad Head, and plan to collaborate with scientists from China, Japan and Korea. They will continuously monitor the dust using radiation detectors, gas chromatographs and mass spectrometers, with further data provided by NOAA aircraft on coastal flights.

Conservationists under fire in the Philippines

Regina Wegner and Quirin Schiermeier

The discovery last year of a new species of two-metre-long lizard focused zoologists' attention on the Visayan Islands in the Philippines, one of the world's biodiversity 'hotspots'. But German and Philippine scientists leading attempts to preserve the rich fauna and flora of this archipelago are facing a campaign of intimidation that they fear will erupt into violence.

Since 1995, the Philippine Endemic Species Conservation Project (PESCP), which is funded by the Frankfurt Zoological Society, has operated a conservation and agroforestry station in the northwest of the island of Panay, some 300 kilometres south of Manila.

The project team is involved in anti-hunting and anti-logging campaigns, and educates local communities in alternative, forest-friendly livelihoods, such as the cultivation of tiger grass, chicken breeding and banana production.

The team's scientific fieldwork includes the release and monitoring of endangered species, such as the Philippine spotted deer (*Cervus alfredi*), the Visayan warty pig (*Sus cebifrons*) and various species of hornbill. Last year, PESCP researchers discovered an arboreal monitor lizard, which they named *Varanus mabitang*. Five new frogs, four snakes and another, smaller lizard species have also been found in recent years.

Recently, however, PESCP members have been exposed to intimidation. "We found cartridge cases and shot hornbills planted in front of the station," says Eberhard Curio, a conservation biologist from the Ruhr University Bochum, who leads the project. "We know that we have enemies among some local families who tried, unsuccessfully, to smuggle poachers into the team. Moreover, some people are suspicious of our campaigns against logging and hunting."

The PESCP team has not yet been attacked, but Curio says the situation on adjacent islands is a cause for serious concern, with scientists having to trek daily to their research sites from secure lodgings.

"On Negros, for example, it has become too dangerous to spend the night in the forests," says Curio. In February, one forest ranger told a local newspaper: "We fled for our lives when we were being attacked by some 100 heavily armed men."

Nonetheless, Curio, who last year was appointed to the first European Union chair for biodiversity in Southeast Asia, will return to the Philippines in August to conduct a radio-telemetry study on released hornbills.

Despite local tensions, biodiversity and conservation programmes in the Philippines are generally well supported by the



Looking up: researchers have discovered several new species in the Philippines, including a monitor lizard (left) in the Visayan Islands.

national and provincial governments. A dedicated biodiversity research unit has also been established at the University of the Philippines Los Baños. "These are very positive signals," says Flint Hobart, who oversees Philippine projects for Conservation International, a Washington-based non-governmental organization.

The Philippines is one of the world's hottest biodiversity hotspots (see *Nature* 403, 853–858; 2000). But it is also the one with the least remaining primary vegetation: only 3% of the original extent is left. The remaining rainforests continue to suffer from illegal logging.



Bush rallies opposition to cloning

Erika Check, Washington

With a vote in the US Senate on banning therapeutic cloning just weeks away, anti-cloning politicians — including President George W. Bush — stepped up their rhetoric last week.

On 10 April, Senator Sam Brownback (Republican, Kansas), who is sponsoring a bill that would outlaw human cloning for any purpose, held a rally for supporters of his legislation. Later that day, many of those supporters moved to the East Room of the White House, where Bush called on the Senate to pass Brownback's bill. The issue is now expected to come to a vote before the end of May.

"Research cloning would contradict the most fundamental principle of medical ethics — that no human life should be exploited or extinguished for the benefit of another," Bush said, to enthusiastic applause.

After Bush's speech, Brownback told reporters that at least 40 senators supported his legislation — 11 votes shy of a majority. However, he added that he was confident of

winning support from enough currently undecided senators to carry the vote.

Supporters of research cloning remain unfazed. "Vote counts differ on this issue," says Kevin Wilson, a spokesman for the Coalition for the Advancement of Medical Research.

But Brownback's opponents were spurred into action by last week's developments. After Bush's statement, senators who support the creation of cloned embryos to provide embryonic stem cells for use in regenerative medicine said they would merge two similar bills, both of which would allow research into therapeutic cloning but not cloning for reproduction.

On the day of Bush's speech, 40 Nobel laureates issued a statement opposing Brownback's bill, saying it would have a "chilling effect" on US science. "Such legal restrictions on scientific investigation would ... send a strong signal to the next generation of researchers that unfettered and responsible scientific investigation is not welcome in the United States," the statement declared.

Europe needs research on primates, say science advisers

Paris Scientists advising the European Commission have reacted to fears that animal-rights groups could pressurize politicians into banning research on non-human primates.

"The European Parliament is currently discussing this," says Albert Osterhaus, a vice-chair of the Scientific Steering Committee and a virologist at Erasmus University in Rotterdam. "A number of members of the parliament are asking for a total ban on non-human primate research."

The 16-member committee, which advises the commission's health directorate, stated on 5 April that the use of non-human primates is necessary for certain areas of biomedical research, such as AIDS and malaria. The committee stressed that primate experiments are needed to ensure the safety of new vaccines before clinical trials, for example.

Former weapons chief walks free after marathon trial

Cape Town Wouter Basson, the former head of South Africa's chemical-weapons programme, was acquitted last week after a

30-month trial on 46 charges — among them murder, fraud and theft.

The charges against Basson included trying to poison guerilla forces during South Africa's occupation of Namibia in the 1980s, and attempting to sell drugs such as ecstasy that he manufactured in government-funded labs.

Judge Willie Hartzenberg said that the case against Basson had not been proved beyond reasonable doubt. He refused to hear evidence of crimes allegedly committed by Basson in Namibia on the grounds that South African courts no longer have jurisdiction there. The director of public prosecutions, Bulelani Ngcuka, has indicated that he will appeal against the decision and that Basson could still be handed over to Namibia for trial.

Ngcuka, who told the Johannesburg



Wouter Basson: acquitted of 46 charges in South Africa — but could yet face trial in Namibia.

Sunday Times that he is "horrified" by the judge's conduct of the case, has said that he will go as far as the constitutional court in order to get a "fair trial".

British brain aims to grab maths million

London A British mathematician claims to have solved a century-old maths puzzle. If correct, he stands to win US\$1 million.

Martin Dunwoody of Southampton University has posted his possible solution to the Poincaré conjecture, an algebraic problem about the properties of multi-dimensional space, on the web. First posed by the French mathematician Henri Poincaré in 1904, the conjecture is one of seven classical mathematical conundrums named as 'Millennium Prize Problems' by the Clay Mathematics Institute in Cambridge, Massachusetts, a non-profit organization that funds and promotes mathematics research. The institute is offering rewards of \$1 million for any correct solutions.

The institute is now reviewing Dunwoody's answer, which could take several months. Other mathematicians are cautious about his chances. "It's a very brave thing to do but the picture is far from clear," says one mathematician familiar with the problem.

♦ www.maths.soton.ac.uk/pure/preprints.phtml

Sydney's latest landmark provokes a mixed reaction

Sydney The construction of a replacement research reactor at Lucas Heights, near Sydney, has been given the green light by Australia's independent safety regulator, despite protests by environmental groups.

Medical and scientific groups welcomed the decision, announced on 5 April by the Australian Radiation Protection and Nuclear Safety Agency, saying that the new facility is needed to replace the existing reactor at Lucas Heights, which has been operating since 1958. The new facility should be operational by 2005. But the move is likely to trigger a fresh wave of protests by environmental activists, who oppose the construction of a nuclear reactor close to residential areas.

Debate over how to deal with waste from the reactor is likely to continue. Spent fuel will be shipped to France for reprocessing, and the resulting intermediate-level waste will be returned to Australia. Sites for storage of this waste have not been identified.

Wellcome Trust scales back rural growth plan

London Two years on from a bitter battle with planning authorities, the Wellcome Trust has resubmitted plans for an extension to its

Genome Campus in Hinxton, Cambridgeshire.

The trust, which is Britain's leading biomedical charity, had its previous application for a 40,000-square-metre extension turned down in 1999 following a campaign by local people. Residents said that the new building would damage the area's rural environment. But the charity is hopeful that its plans will be approved this time around — the proposed extension would cover 27,000 square metres, which is in line with the scale recommended by the government when it rejected the trust's previous application.

Space has been saved by dropping plans to house established biomedical companies on the site. But the extension will still include space for genomics start-up companies and further academic research facilities for the Sanger Institute, which played a major role in sequencing the human genome.



Wellcome compromise: plans to enlarge the charity's Cambridgeshire campus have been revised.

UK environment agency falls foul of new accounting rule

London Researchers working for Britain's Natural Environment Research Council (NERC) could face funding delays and budget cuts after an accounting mix-up left the organization with a £14-million (US\$20-million) hole in its finances.

NERC was caught out by new government accounting rules, which separate capital and resource budgets that could previously be used to balance each other. Officials at the government's Office for Science and Technology have written to colleagues at the Treasury, urging them to exempt NERC from new regulations, for which NERC says it was not given sufficient time to prepare.

If it is not exempted, NERC will have to save the money from its £200-million annual budget by 2004, meaning that grants could be delayed, staff at NERC institutes not replaced, and science budgets cut.

Break-bone fever

The dengue virus exacts a devastating and growing toll on public health in the tropics, yet remains little studied. Tom Clarke talks to the scientists who are intent on raising dengue's profile.

WHO/TDR/STAMMERS

"You don't die from it, but you wish you could," says Duane Gubler. He should know — he's been stricken by dengue fever three times. Gubler's first infection consigned him to bed for more than a week with a raging temperature and the agonizing limb pains that have earned the disease its sobriquet 'break-bone fever'. In addition, some sufferers lose hair and develop a measles-like rash, bleeding gums and depression that can last for weeks.

They're the lucky ones. In a small percentage of cases, mostly in infants, the infection causes dengue haemorrhagic fever (DHF)—which is a killer. Its victims vomit and pass blood in their faeces and urine as their capillaries leak fluid. If spotted early, DHF responds to intensive hospital treatment. But if it isn't caught, mortality can reach 15%.

The dengue virus is spread by the mos-



Smoking out the enemy: a pest controller in Havana, Cuba, fumigates a house against *Aedes aegypti* (left), the mosquito that transmits the dengue virus to humans.

quito *Aedes aegypti*. For Gubler, the disease is an occupational hazard — he has spent his career studying dengue in the field, and now directs the vector-borne diseases division of the US Centers for Disease Control and Prevention, based at the National Center for Infectious Diseases in Fort Collins, Colorado.

Growing concern

Once a sporadic illness, dengue is on the rise. Epidemics are now regular in southeast Asia, India, the western Pacific and much of South America (see map, opposite). Outbreaks are also getting bigger and more serious, with a higher proportion of DHF cases. The World Health Organization (WHO) receives reports of about 500,000 dengue fever cases each year, but estimates that as many as 50 million people are infected annually.

Yet the disease is neglected. Almost entirely absent in the developed world and difficult to study, dengue has received little attention — and a fraction of the research funding devoted to better-known tropical diseases. "It's malaria's poor cousin," says Gubler.

Although dengue doesn't rival malaria as a killer, the sheer number of people infected and the fever's debilitating nature mean that it has an enormous economic impact. Measured in units devised by the World Bank called disability-adjusted life years, which

quantify disruption to quality of life and economic productivity, dengue's burden on some societies is comparable to that of HIV, tuberculosis or hepatitis. In South America, its impact rivals that of malaria, and in southeast Asia, it is public-health enemy number one¹.

"It's an orphan disease, but one that's increasingly screaming at us," says Mike Nathan, a dengue specialist at WHO's headquarters in Geneva. He, Gubler and other experts are now trying to push dengue up the public-health agenda — with some success. In January, WHO's executive board drafted a resolution urging international agencies and national governments to spend more on studying and tackling dengue. This will be considered by the World Health Assembly, WHO's governing body, next month.

This effort to raise dengue's profile is partly due to renewed hope that it will be possible to produce a vaccine, which would bolster efforts in mosquito control and disease surveillance. But this will only happen, say dengue experts, if scientists rise to the challenge and are supported by adequate funds.

How did the situation get so bad? Demographic changes — particularly urbanization — are largely to blame. *Ae. aegypti* prefers to feed on human blood, and can spread the dengue virus most effectively when people are living cheek-by-jowl². Mas-

JOSE GOITIA/AP

GERALD ANKER/CORBIS



Huge troop deployments in the Second World War sowed the seeds of Asia's dengue problem.

sive troop deployments and refugee movements in the Second World War established southeast Asia as the world's dengue hotbed. But in the years after the war, efforts to combat yellow fever and malaria held the disease in check by ridding the region of *Ae. aegypti*.

Dengue's comeback was triggered by the arrival in South America of widespread vaccination for yellow fever in the 1950s and the scaling back of a worldwide anti-malaria initiative in the 1970s. Mosquito-control efforts slackened, as urbanization continued to gather pace³. Over the past five decades, the disease's incidence has grown 30-fold⁴. But its tendency to lie dormant in any given region before exploding into a severe outbreak has contributed to dengue's neglect by public-health experts. Cuba and Brazil have this year been hit by severe epidemics. But in countries experiencing a lull, it is easy for health authorities to give dengue a low priority. "With tight budgets, the disease gets pushed into a corner," says Nathan.

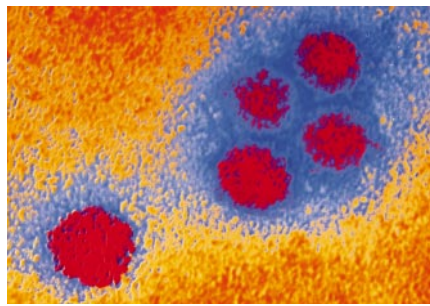
Messy business

Any effort to control dengue will involve spraying with insecticides, and outlawing of open water containers and litter such as discarded tyres — which can collect water and serve as perfect nurseries for aquatic mosquito larvae. But although mosquito control works reasonably well, it is "a little bit messy", says Nathan. As such, it can be hard to sell to public-health authorities. A vaccine, however, would be a different matter — more likely to gain a high profile and financial backing. "To many people's eyes it's a tool that's neat and clean," says Nathan.

Work began on dengue vaccines in the 1970s, but progress has been slow. Part of the problem is that — unlike yellow fever — dengue is caused not by a single virus but by four distinct viral 'serotypes'. Each is equally infectious and pathogenic. And the catch is that once a person's immune system has fought off and memorized an infection by one dengue serotype, a secondary infection with a different serotype is thought to be the biggest single risk factor for developing DHF.

The reasons for this are not well understood. Different dengue serotypes are seen as similar by the immune system — when a new serotype infects someone who has had a previous bout of dengue, it is recognized. But although antibodies bind to the new virus, they do not do so as effectively as they would to the serotype they encountered previously. So when immune cells called macrophages arrive on the scene to ingest the virus-antibody complexes, the virus remains infectious. And unfortunately, macrophages are the very cells that dengue prefers to infect.

As a result, exposure to a second dengue serotype can cause higher levels of infection than occurred the first time around^{5,6}.



Virus and vector: left, dengue virus particles; right, Brazilian children examine *Aedes aegypti* larvae.



Researchers led by Francis Ennis, an immunopathologist at the University of Massachusetts Medical School in Boston, have found that this has knock-on effects on another arm of the immune system: the 'killer' T cells that attack virus-infected cells. When exposed to a burgeoning infection with a second dengue serotype, the T cells can overreact. "It's like an immunological explosion," says Ennis. The killer cells start producing excessive quantities of cytokines⁷, molecules that at normal concentrations help to coordinate immune responses. In excess, they can cause other cells, and capillaries, to leak fluid. This, Ennis suspects, leads to DHF⁸.

The phenomenon is worst in very young children with naive immune systems. They may receive antibodies against one serotype in their mother's milk and then become infected by a mosquito carrying a different one. So any reliable dengue vaccine must provide complete protection against not just one serotype but all four, and be safe for babies. Otherwise, vaccination in an endemic area could cause a rise in DHF incidence, even if it reduces the total number of dengue cases.

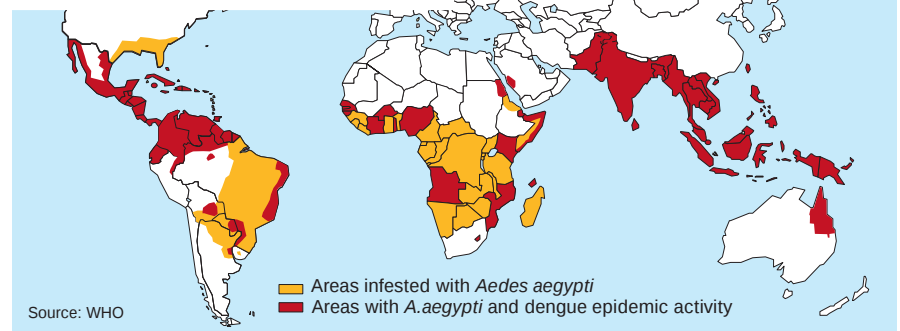
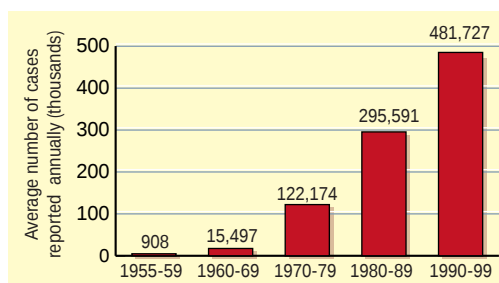
To make matters worse, says Alan Rothman, who works in Ennis's team, there is no

good animal model with which to study dengue and DHF. "Mice get infected, but they don't get sick in any way," he says. Although monkeys produce antibodies against the virus and sometimes develop a fever, they never get the haemorrhagic form. So research into dengue vaccines involves a largely blind leap from the test tube to clinical trials.

Promising prospects

Despite these obstacles, dengue vaccines are now edging towards clinical use. Two of them, one developed at Mahidol University in Bangkok, Thailand⁹, and another at the Walter Reed Army Institute of Research in Silver Spring, Maryland¹⁰, are based on living, weakened dengue viruses, selected by growing all four serotypes through multiple generations in cultured cells. The Bangkok vaccine is now licensed to Aventis Pasteur of Lyon, France. Safety and efficacy trials are ongoing and show that the vaccine is safe and protects up to 90% of recipients from all four serotypes. The Walter Reed vaccine, licensed to GlaxoSmithKline Biologicals in Rixensart, Belgium, will begin safety and efficacy trials in infants later this year.

Other candidates are on the way. The National Institute for Allergy and Infectious Diseases in Bethesda, Maryland, is working on a dengue serotype in which virulence genes are deleted¹¹ — early indications are that this is safe and produces an antibody response. The next step is to insert genes that encode the



Global trend: the wide distribution of *Aedes aegypti* — largely due to lapses in efforts to control mosquito populations — means that dengue fever can strike throughout the tropics.

▶ protein coat of the other three serotypes. Similarly, Gubler's team has genetically modified one of the weakened Bangkok viruses by adding coat-protein genes from the other three serotypes¹². This vaccine has stimulated antibody production in rhesus monkeys.

One-shot deal

Another promising candidate is based on a modified form of the successful yellow fever vaccine — a live, weakened yellow fever virus carrying genes for the dengue protein coat, developed by Acambis of Cambridge, UK. Unlike the Bangkok and Walter Reed vaccines, which require two rounds of vaccination to provide immunity, the Acambis vaccine should require just one shot.

Encouragingly, Acambis has used the same approach to develop a vaccine against Japanese encephalitis, which is caused by a virus related to dengue, but with only one serotype. This vaccine is performing well in clinical trials¹³. Although the company has so far added genes from only one serotype to its dengue vaccine, it intends to add those from the other three. "We still need to show that immunity is lasting and persists for all four serotypes," says Tom Monath, head of vaccine development at Acambis, based at its US branch in Cambridge, Massachusetts.

Other approaches are at earlier stages of development. The US Naval Medical Research Center in Silver Spring and the biotech firm Maxisgen of Redwood City, California, for instance, are both developing a vaccine consisting of genes for dengue coat proteins, packaged with other DNA that promotes the recognition of these proteins by the immune system. Each has been tested in animals, and human safety trials are planned for this year. Such 'naked' DNA vaccines¹⁴ have the advantage of being completely non-infectious — whereas live, weakened vaccines can potentially revert to a disease-causing form.



Trail of blood: Thomas Scott is using DNA analysis to shed light on mosquitoes' activities.



Breeding ground: open water containers serve as nurseries for the dengue virus's mosquito vector.

Given that dengue mostly afflicts poor countries, a vaccine is unlikely to be a major moneyspinner. So experts argue that public funds may be needed to bring vaccines to the people who need them. To focus fundraising and vaccine-development efforts, and to provide a forum to discuss the merits of the various candidates, the New York-based Rockefeller Foundation intends to set up a dengue vaccine programme at the International Vaccine Institute in Seoul, South Korea — which was established to promote the development of vaccines for diseases of the developing world. "This is a critical effort," claims Scott Halstead, a dengue researcher and the Rockefeller's associate director of health sciences.

Know your enemy

But even if Halstead and others successfully invigorate vaccine development, it will be years before vaccination against dengue is routine. In the meantime, a few researchers are getting intimate with dengue and its mosquito vector. They hope to understand exactly how the disease spreads, why it waxes and wanes — and perhaps to discover new ways to avoid or prepare for epidemics.

Working near Mae Sot in western Thailand, Thomas Scott of the University of California, Davis, has found that a female *Ae. aegypti* may need five blood meals to reproduce successfully, compared with the single blood meal required by most other mosquito vectors¹⁵. This finding, confirmed by individually marking female mosquitoes and recapturing them around people's homes, helps to explain why dengue can persist even when very few mosquitoes are present. It also shows that just a few infected mosquitoes or people are needed to kickstart an outbreak. "It's not good news for those trying to eradicate dengue by killing mosquitoes," says Scott.

Right now, Scott is developing techniques to evaluate the risk of dengue outbreaks in specific communities. Using DNA fingerprinting, he has been able to match the blood in captured mosquitoes' stomachs to DNA samples from villagers. He is now using this technique to discover which segment of the population mosquitoes bite, how frequently, and how far they roam in search of a meal¹⁶.

In another project in Iquitos, Peru, Scott's team has begun to relate the dynamics of dengue transmission to the density of mosquito populations. They are building on previous research in which Halstead's team took blood samples from 1,300 schoolchildren and screened them for antibodies against dengue. They then monitored the arrival and subsequent spread a new serotype of dengue in this isolated Amazonian city, which previously has been host to only one serotype¹⁷.

Basic research such as this, Scott hopes, may provide clues that will allow public-health officials to predict an imminent outbreak. "Our hope is that we can identify some fundamental principles that can be tested elsewhere," he says. Predicting an epidemic before it happens might not help to stop it, but it could help health departments prepare to treat the youngest and most susceptible victims, and so prevent unnecessary deaths.

After decades of neglect, Gubler is confident that dengue research and efforts to develop an effective vaccine will soon start getting the attention that he and others believe they deserve. "The other option is to sit around and watch thousands die every year," he says. ■

Tom Clarke works in Nature's science writing team.

1. Gubler, D. J. & Meltzer, M. *Adv. Virus Res.* **53**, 36–70 (1999).
2. Scott, T. W., Naksathit, A., Day, J. E., Kittayapong, P. & Edman, J. D. *Am. J. Trop. Med. Hyg.* **57**, 235–239 (1997).
3. Halstead, S. B. in *Dengue and Dengue Hemorrhagic Fever* (eds Gubler, D. J. & Kuno, G.) 22–44 (CABI, Wallingford, UK, 1997).
4. Gubler, D. J. *Trends Microbiol.* **10**, 100–103 (2002).
5. Vaughn, D. W. *et al. J. Infect. Dis.* **181**, 2–9 (2000).
6. Halstead, S. B. *Rev. Infect. Dis.* **11**, Suppl. 4, S830–S839 (1989).
7. Rothman, A. L. & Ennis, F. A. *Virology* **257**, 1–6 (1999).
8. Gagnon, S. J. *et al. J. Med. Virol.* **67**, 41–46 (2002).
9. Bhamarapravati, N. & Sutee, Y. *Vaccine* **18**, 44–47 (2000).
10. Eckels, K. H., Harrison, V. R., Summers, P. L. & Russell, P. K. *Infect. Immun.* **27**, 175–180 (1980).
11. Durbin, A. P. *et al. Am. J. Trop. Med. Hyg.* **65**, 405–413 (2001).
12. Huang, C. Y. *et al. J. Virol.* **74**, 3020–3028 (2002).
13. Monath, T. P. *et al. Vaccine* **20**, 1004–1018 (2002).
14. Chang, G. J., Davis, B. S., Hunt, A. R., Holmes, D. A. & Kuno, G. *Ann. NY Acad. Sci.* **951**, 272–285 (2001).
15. Scott, T. W. *et al. J. Med. Entomol.* **37**, 89–101 (2000).
16. Chow-Schaffer, E., Hawley, W., Sina, B., DeBenedictis, J. & Scott, T. W. *J. Med. Entomol.* **37**, 492–502 (2000).
17. Watts, D. M. *et al. Lancet* **354**, 1431–1434 (1999).

WHO dengue factsheet

▶ www.who.int/inf-fs/en/fact117.html

Draft resolution for World Health Assembly

▶ www.who.int/gb/EB_WHA/PDF/EB109/eeb109r4.pdf



J. ALEXANDER/GETTY

DIY, Kiev style

Many Ukrainian research institutes went to the wall when the Soviet Union collapsed. Quirin Schiermeier finds out how home-grown talent — and home-made equipment — have helped one centre to buck the trend.

The Soviet-era tower block that dominates Bogomoletz Street is a prestigious address in the Ukrainian capital of Kiev. US investment bankers occupy one office. From another, the television channel OTV broadcasts music videos to Ukraine's youth. And at the top of the building, a spacious, well-equipped animal house is home to hundreds of mice, rabbits and guinea pigs.

The animal facilities belong to the A. A. Bogomoletz Institute of Physiology, a research centre whose reputation has survived the turmoil that followed the collapse of the Soviet Union at the start of 1990s. Once the breadbasket of the communist

superpower, Ukraine was left with just crumbs — and scientists suffered alongside everyone else. A quarter of the Soviet Union's immense research resources once flowed into Ukraine. Now, annual government funding for the country's 150,000 researchers and engineers has dipped below US\$100 million — a fraction of western European science budgets.

Despite this, the Bogomoletz Institute continues to produce quality research — work that neuroscientist Erwin Neher, joint-winner of the 1991 Nobel Prize in Physiology or Medicine, describes as “world-class”.

Channel vision

The institute has an impressive history. Located in the heart of Kiev, it was created in 1953 when the Ukrainian Academy of Sciences' Institute of Clinical Physiology was merged with the Institute of Experimental Biology and Pathology, part of the Ukrainian health ministry. Thanks largely to the work of Platon Kostyuk, who joined the institute in 1958 and today serves as its director at the age of 78, the Bogomoletz became the Soviet Union's premier centre for neurophysiology.

During a 1961 sabbatical at John Eccles' laboratory at the Australian National Univer-



High riser: despite financial worries, Kiev's Bogomoletz Institute (above) is world-renowned.

sity in Canberra, Kostyuk learnt how to use very fine electrodes to probe the workings of brain cells. He introduced the technique to the Bogomoletz Institute on his return — the first time the procedure had been used in the Soviet Union. Subsequently, Kostyuk and his colleagues made big contributions to our understanding of ion channels in nerve cell membranes. They were the first, for example, to show that ion channels can be separated into low- and high-voltage types. By the time Kostyuk became director in 1966, the institute had developed an international reputation.

Kostyuk and his colleagues went on to develop an early variant of the patch-clamp technique. This method, which revolutionized cell physiology, involves monitoring the opening and closing of ion channels by placing a fine pipette filled with a conducting solution over a patch of membrane. The technique was eventually perfected in the late 1970s by Neher and Bert Sakmann at the Max



Self-sufficient: building their own equipment has allowed Bogomoletz researchers to keep working.

Q. SCHIERMEIER

Q. SCHIERMEIER

► Planck Institute for Biophysical Chemistry in Göttingen, Germany, for which the pair earned their Nobel.

When the Soviet Union unravelled, it became clear that the Bogomoletz Institute faced a difficult future. Withdrawal of Soviet funding has devastated Ukraine's research base, and like most of the country's 1,500 research centres, the Bogomoletz experienced hardship. But an influx of foreign funding and a knack for building their own equipment have allowed its best researchers to remain internationally competitive.

Speech therapy

The institute's education programme, created by Kostyuk, is at the heart of its success. Would-be brain researchers undergo a tough three-year training in mathematics, physics and chemistry, conducted in both English and Russian, before they join a group at the Bogomoletz to specialize in physiology. In the Soviet era, this training was done at the Moscow Institute of Physics and Technology; today, the students go to the Physical Engineering Centre in Kiev.

At weekly seminars, Kostyuk keeps alive the relaxed international flair and scientific openness he captured in Australia. Research students gather each Friday in a draughty lecture hall to discuss their work with Kostyuk and visiting academics — talks that Kostyuk insists are held in English.

But some researchers inevitably chose to move on. Olga Garaschuk left the institute in 1992 to take up a position at the Ludwig-Maximilians University of Munich. She says her work on calcium signalling in developing brains would be impossible without the advanced optical imaging technology that is available in her lab. The four microscopes in Garaschuk's lab each cost at least \$200,000 — far more than the Bogomoletz can afford.

But a lack of resources does not always prevent Bogomoletz researchers from getting the equipment they need. Nickolai Veselovsky uses microelectrode techniques to study how calcium ions trigger brain cells to release chemical signals. His group owns

some modern purpose-built, German-made microscopes, each of which cost \$20,000. But other equipment, such as the micromanipulators used to position the electrodes, has been built in-house.

Some researchers have even cannibalized Soviet weapons. Elena Lukyanetz, a young biophysicist, images the intracellular ions that prompt brain cells to release neurotransmitters — chemical signals used to communicate between nerve cells. She uses a device built from optical equipment salvaged from a dismantled SS-20 missile.

Parts of the institute may have an air of do-it-yourself, but Veselovsky claims that his equipment allows state-of-the-art research on nerve transmission — a view shared by Neher, who coordinates some projects at the institute under a UNESCO-funded scheme. "After all," says Veselovsky, "constructing your own devices has a strong tradition in classic physiology."

Joint effort

The institute also exploits the connections gained when its researchers move to foreign centres. Around 130 former Bogomoletz scientists are working in European and North American institutes. Alexej Verkhratsky, currently at the University of Manchester, UK, is one. He is collaborating with Veselovsky, one of seven Bogomoletz researchers who between them have received \$1 million from the Wellcome Trust, Britain's leading biomedical research charity.

Such links are helping the Bogomoletz through difficult times, but the institute has long been outward-looking. It has, for example, hosted a regular international meeting, organized in conjunction with physiological societies in Spain, Italy, Germany, Britain and the United States, since the 1970s. "It was our international reputation which helped us survive when we were suddenly forced to look for money ourselves," says Kostyuk. Today, grants from bodies such as the NATO Science Programme, which supports scientific collaborations between countries in the Middle East, eastern Europe and NATO member states, are an elixir of life.

Lukyanetz's group has, for example, recently received an \$80,000 grant from INTAS, an independent body supported by the European Union to promote East-West scientific cooperation. She is using the funds to study how packets of neurotransmitters are moved around within cells, and how proteins found at the junctions between cells influence this movement. The collaborative project, coordinated by Neher, involves groups from Germany and Britain.

Kostyuk has also found other means of bringing money into the institute, such as the subletting schemes that have brought OTV and investment bankers into the building. Thanks to its privileged location in the city centre, the institute can charge rents equal to



With his book finished, Oleg Krishtal hopes to market his patch-clamp robot for screening drugs.

those of other European capitals.

Contract research for western pharmaceutical companies is also becoming a substantial source of additional income. "We are able to deliver complete pharmacological profiles of substances within only three days," says Oleg Krishtal, deputy director of the institute and head of cellular membrane research. Some researchers familiar with the Bogomoletz say that this contract work includes animal experiments. Kostyuk declines to comment, but concedes that it is much easier to get permissions under Ukraine's relaxed animal-protection laws than in the West.

The institute also hopes to start its own companies. Krishtal is planning to use a homemade patch-clamp robot to examine the pharmaceutical potential of toxins in the venoms of cone shells, a type of mollusc. Some of these toxins inhibit one of the major pain receptors found in neurons. He hopes to launch a start-up company in collaboration with Andrew Miller, a director of the new Genetic Therapies Centre at Imperial College, London, and Leonid Moroz, a neuroscientist at the University of Florida in St Augustine.

Staff at the Bogomoletz Institute are glad that Krishtal is once again dedicating all of his energy to research. Described by Neher as a visionary scientist, he has recently finished his book *To The Singing of Birds*, a poetic and philosophical examination of consciousness and evolution that has been widely acclaimed in the Russian and Ukrainian media. No date has been set for Kostyuk's retirement, but many researchers at the Bogomoletz hope that Krishtal will take over when he does finally step down.

Whoever succeeds Kostyuk will have a tough act to follow. Ukraine's politics and economy may be more settled than in the immediate aftermath of the Soviet collapse, but the institute's funding is still far from certain. Its researchers are proud to work in eastern Europe's most prestigious neuroscience centre. But Kostyuk's successor will have to work hard if they are to keep their crown. ■

Quirin Schiermeier is *Nature's* German correspondent.

♦ www.biph.kiev.ua



Father figure: Platon Kostyuk's scientific vision has helped to maintain the institute's success.

Work on weapons adds to public distrust of science

Academics need to remain independent and remember their ethical obligations.

Sir—The partial transparency concerning the technology needed to retain Britain's nuclear weapons, represented by the Feature "Science of nuclear warheads" (*Nature* **415**, 853–857; 2002) by Keith O'Nions *et al.*, is to be welcomed. Yet it is disappointing that the authors do not discuss the alternative of decommissioning and destroying such weapons — required if Britain is to adhere to international law and to the agreements it has signed, such as the Nuclear Non-Proliferation Treaty.

The apparent subordination of science to government is one cause of current public distrust of scientists and of student reluctance to embark upon scientific careers. Views such as those in the article concerning nuclear-warhead science can only add to this alienation.

O'Nions *et al.* call for new appointees to develop "a deeper theoretical and

experimental understanding" of the science needed to underwrite "the safety and performance of the ageing Trident stockpile". But they do not explain what "performance" means in this context, and could perhaps agree that in a slightly modified political climate, safety could be assured by dismantling and destroying warheads.

The problems of doing this, and the verification needed, may well require the temporary involvement of high-calibre scientists and engineers. Whether we can or should hire them to maintain the threats is doubtful, and few able students interested in fundamental science are likely to see the latter as a desirable career.

The US National Ignition Facility and the supercomputing devices discussed in the Feature are means by which weapons of mass destruction can be developed and partially tested without breaching the letter

of the Comprehensive Test Ban Treaty. O'Nions *et al.* describe a retention programme that, they say, would not only recruit high-calibre staff but "work with British academics and industry" to retain indefinitely "the reliability and safety of Britain's independent nuclear warhead".

This ignores the need for the "British academic community" (if there is such a thing) to stay independent of government pressure; to consider the ethical dimension of their collaborations; to continue open international links, including with the former Soviet Union and politically unstable parts of the developing world; and to advise students on careers with both intellectual challenge and moral probity.

Peter Nicholls

Chair, Abolition 2000 UK, and Department of Biological Sciences, University of Essex, Colchester, Essex CO4 3SQ, UK

Mutation not universally linked with diabetes

Sir—Pierre Maechler and Claes B. Wollheim¹ in their interesting Insight Review Article on mitochondrial function in normal and diabetic β -cells show diabetes-associated mutations in the human mitochondrial genome in their Fig. 1. One of these mutations is G3316A in the *ND1* gene (references cited in ref. 1).

Until recently, this mutation had been reported only in a Japanese population. Since then, other independent studies^{2,3} have not found this association between G3316A and diabetes in Chinese people or in another Japanese population⁴.

Because about 4% of China's 1.2 billion people carry this mutation², it is important that the polymorphism G3316A is not regarded as diabetes-associated outside the Japanese population reported in ref. 1.

Ching-Wan Lam

Department of Chemical Pathology, The Chinese University of Hong Kong, Hong Kong, China

1. Maechler, P. & Wollheim, C. B. *Nature* **414**, 807–812 (2001).
2. Lam, C. W. *et al.* *J. Med. Genet.* **38**, E10 (2001).
3. Ji, L., Hou, X. & Han, X. *Diabetes Res. Clin. Pract.* **54**, Suppl. 2, S35–S38 (2001).
4. Matsuura, N. *et al.* *J. Pediatr. Endocrinol. Metab.* **12**, 27–30 (1999).

Single-locus studies

Sir—Neil Risch *et al.* in Correspondence¹ state that our paper² on the genetic relatedness of Palestinians and Jews lacked scientific merit because its conclusions are based on data reported for a single-locus

genetic marker (HLA-DRB1). Although the use of single-locus markers can lead to misleading results, single-locus studies, whether using HLA or other markers, are common in this field and are regularly published in the specialist literature.

In papers reporting data on a single locus, it is important not to take anomalous results at face value but to interpret them in the light of other types of data, such as historical, anthropological and linguistic data, as well as testing them using other genetic markers (see, for example, ref. 3). As we stated in ref. 2, we are currently investigating the populations reported in our paper using other markers.

Antonio Arnaiz-Villena, Eduardo Gomez-Casado, Jorge Martinez-Laso

Department of Immunology (Microbiology), Faculty of Medicine, Universidad Complutense, 28040 Madrid, Spain
e-mail: aav@efd.net

1. Risch, N., Piazza, A. & Cavalli-Sforza, L. L. *Nature* **415**, 115 (2002).
2. Arnaiz-Villena, A. *et al.* *Hum. Immunol.* **62**, 889–900 (2001).
3. Dork, T. *et al.* *Am. J. Hum. Genet.* **63**, 656–662 (1998).

Call for Elan to publish Alzheimer's trial details

Sir—One potential route for therapy for Alzheimer's disease involves the immunization of patients against the amyloid- β peptide, the major proteinaceous component that characterizes plaques in the brains of patients with this disease. Resulting from earlier research on mice, published in *Nature*^{1–3},

Elan Pharmaceuticals recently began clinical trials of a vaccine, AN-1792, based on this approach. In January, four patients in this trial developed inflammation of the central nervous system, and Elan suspended the trial pending further investigation⁴. By March, the number of patients in this trial who had symptoms resembling encephalitis and meningitis was reported to have increased to 15, and Elan has permanently withdrawn the AN-1792 vaccine from human trials⁵.

It is not clear precisely what caused the inflammation in these patients, as Elan has released few details of their medical condition. Elan is under no obligation to report the results of the AN-1792 trials, and may well not wish to lose its perceived scientific or commercial advantage.

Nonetheless, detailed information of the conditions developed by these patients is of profound interest to the scientific community as well as society at large. We therefore urge Elan to publish details from their study as quickly as possible, so that any foreseeable side-effects can be avoided in future trials.

Glenda M. Bishop*, Stephen R. Robinson†, Mark A. Smith*, George Perry* & Craig S. Atwood*

**Institute of Pathology, Case Western Reserve University, Cleveland, Ohio 44106, USA*

†School of Psychology, Psychiatry and Psychological Medicine, Monash University, Clayton, Victoria 3800, Australia

1. Schenk, D. *et al.* *Nature* **400**, 173–177 (1999).
2. Janus, C. *et al.* *Nature* **408**, 979–982 (2000).
3. Morgan, D. *et al.* *Nature* **408**, 982–985 (2000).
4. Check, E. *Nature* **415**, 462 (2002).
5. Weiss, R. *Washington Post* A03 (2 March 2002)

The story of an ecosystem

A ten-year study of a Canadian forest shows the way ahead for ecology.

Ecosystem Dynamics of the Boreal Forest: the Kluane Project
 edited by Charles J. Krebs, Stan Boutin & Rudy Boonstra
Oxford University Press: 2001. 544 pp. \$95.
 Nils Chr. Stenseth

Just as every picture tells a story, so too does every scientific project. This book is the story of the Kluane Boreal Forest Ecosystem Project, told brilliantly by the researchers who toiled for ten years from 1986 to uncover the drama and complexity of this Canadian ecosystem. Whoever wants to learn about the boreal forest of North America, one of the largest ecosystems of the world, and the stage on which a classic ecological drama — the ten-year population cycle of the snowshoe hare — is played out repeatedly, should definitely read this book. Equally important, it should be read for what it tells us about how to bring ecology into the twenty-first century: namely, to integrate population ecology into ecosystem ecology. The large-scale ecosystem study presented in this book represents a major step in that direction.

The book is illustrated with excellent drawings by Carol Stefan and accompanied by a CD-ROM put together by Alice Kenney. Although the book reports on highly integrated teamwork, where many scientists have played a significant part both in the planning and the execution of the project, it also tells the story of the Arctic endeavours of one man — Charles Krebs. Inspired as a young boy by stories of Arctic pioneers such as Roald Amundsen and Fridtjof Nansen, he began work for his PhD on the lemming cycle in the late 1950s at Baker Lake, near Hudson Bay in Arctic Canada. As professor of zoology at the University of British Columbia, he organized the Kluane project in the western part of northern Canada. No wonder that the other members of the project express their “heartfelt thanks to Charles J. Krebs. Charley was the vital force behind this research collaboration. In no small measure, the successes of the study were related to his vision, his leadership, and his ability to keep us all focused on the task for over a decade of field research and writing.”

The Kluane project was carried out in the southwestern Yukon. Nine faculty members from three Canadian universities and 26 graduate students joined with 75 summer assistants and 18 technicians to expend 157 person-years of effort. This team gained considerable benefit from many studies that



Hare today, gone tomorrow? Populations of the snowshoe hare rise and fall over a ten-year cycle.

had been carried out since the 1950s at the Arctic Institute of North America's Kluane Research Station.

The Kluane project set out to tackle one of the key — and longstanding — questions in the field of ecology: what regulates the dynamics of species within an ecosystem such as the boreal forest around Kluane? Are species regulated from below (through primary productivity) or from above (through predators), or a combination thereof?

The project found that much of the boreal forest vertebrate community is regulated by top-down control: several predatory species act as a guild within which each species operates in a compensatory manner. Populations of guild members are influenced to a greater or lesser extent by the population numbers of the major prey species — the snowshoe hare. Thus, some of these predator species are potentially redundant. But some species (such as the red squirrel) dance to a different drummer, being little influenced by the dramatic and regular fluctuations in snowshoe-hare populations, which are known to affect many other species, such as the lynx. Key factors in the ecosystem, such as nutrients, strongly influence plant growth, but their effects do not cascade upwards to affect predator populations.

As an important by-product, the project has given us a detailed description of the food web of the boreal forest of North America and, equally important, a unique database that will serve as a marvellous platform for ecologists in the years to come.

The Kluane project was not able to

address the issue of spatial synchrony in the ten-year cycle — why hare populations across most of the 5 million square kilometres fluctuate together. However, this book provides a good basis for further theoretical work on such spatial issues. The authors recommend that similar large-scale projects are set up in other parts of the Canadian boreal forest. Indeed, that would provide the basis not only for extrapolating the results obtained to regions outside the Kluane Lake district, but also for addressing some of the spatial issues mentioned above.

Although the book does not address the issue of the ecological impact of global change, the work will serve as a benchmark from which to assess such change. It puts us in a much better position to address questions relating to which species — and species interactions — might be affected, and the possible resulting change in landscape. Indeed, to address the issue of the possible effects of climate change, monitoring of this boreal-forest community (which has now been studied for more than 25 years) should be continued. By so doing, we would be able to integrate observational and experimental findings within a combined modelling effort — just what is needed to address the ecology of climate change.

The Kluane project was not without its practical challenges. The fact that such a large-scale project has been completed, however, should encourage other ecologists to embark on similar large-scale endeavours. Indeed, I am convinced that it is through such large-scale projects that we

will be able to advance our broader ecological understanding.

This type of project is also exactly what is needed to enable ecologists to help politicians manage the biological diversity of the Earth when faced with a growing population, and the resulting increase in demand for resources. All such demands must ultimately be met from our natural resources. Research-funding agencies, and hence politicians, must realize that it is not enough to have had one Kluane project. We need many, so that we can compare the dynamics of ecosystems under different settings. Before the Kluane Project we had no role model — now we have one. ■

Nils Chr. Steneth is in the Division of Zoology, Department of Biology, University of Oslo, PO Box 1050 Blindern, N-0316 Oslo, Norway.

A power that's clean and bright

Clean Electricity from Photovoltaics

edited by Mary D. Archer & Robert Hill
Imperial College Press: 2001. 868 pp. £82, \$120

Richard Corkish

This comprehensive summary of photovoltaics (solar cells) comes amid growing concern over global warming and energy security. The Sun reliably delivers each year more than 10,000 times the global consumption of commercial energy. All the world's primary energy requirements could be met by solar cells taking up an area less than 0.25% of that presently under crops and pasture.

Much of this book, the first volume in a series on the photoconversion of solar energy, is concerned with detailed descriptions of the various technologies by which solar cells are made. Most of the important ones have a chapter devoted to them, written by experts in the field. The terrestrial market is currently dominated by cells made from wafers of crystalline silicon. This situation has been expected to change for decades but, so far, silicon remains top of the heap, partly because of its ability to piggy-back on developments in the electronics industry.

There are at least two divergent paths for the development of other cell types. The first is to make them more cheaply at the expense of efficiency. Amorphous silicon, cadmium telluride, copper indium gallium diselenide and thin-film polycrystalline silicon are all in production now or soon will be. Another technology, dye-sensitized cells, is not mentioned here but will be included in the third volume in the series. Organic photovoltaics are probably the ultimate low-cost approach, and their progress is detailed here.

The second path is the pursuit of very high efficiencies, accepting higher costs and using the cells in situations where the extra efficiency pays for itself in space and in concentrators. The most advanced of these are actually stacks of subcells, each made from a different material and each accessing a different part of the solar spectrum. The photovoltaics research community has quite a few other concepts for potentially obtaining greater efficiency, and one of these, the quantum-well solar cell, is fully described here. There are also several recent proposals, still in the very early stages of research, aimed at combining the best aspects of the low-cost and high-efficiency paths, but they are not included here. Each of the above cell technologies is covered in great depth, with the exception of polycrystalline silicon, which, I think, deserved its own chapter.

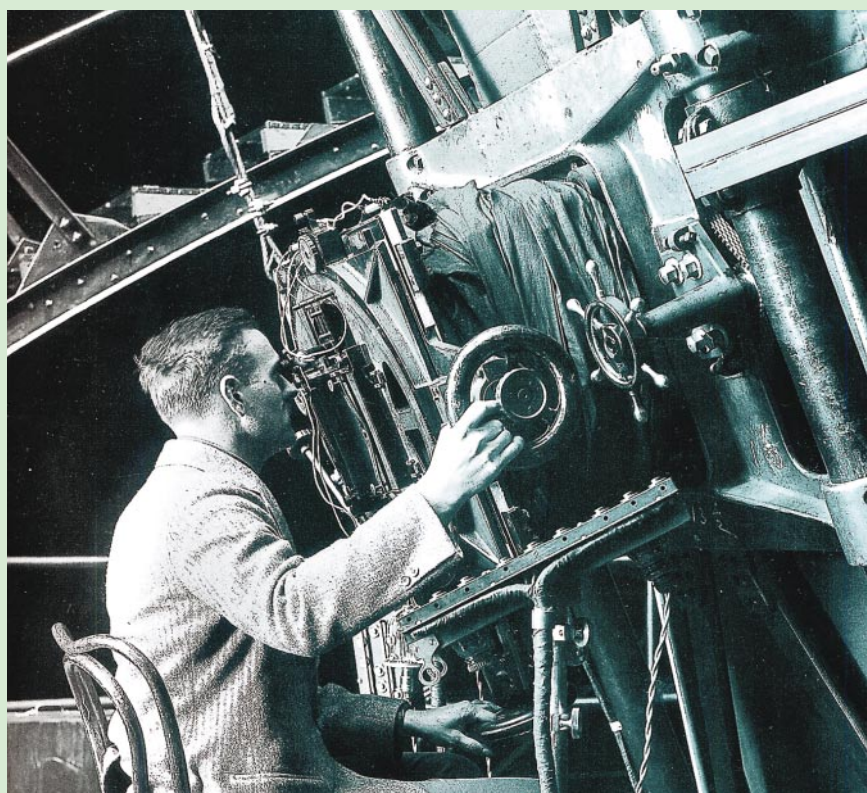
The rest of the book is concerned with how we can, do and should use solar cells. There is one chapter on space applications, but the remainder are concerned with powering human needs on Earth. Energy storage and system aspects are well documented in separate chapters, and the final three outline the photovoltaic business situation, economics, policy proposals and future prospects,

with attention given to applications in both developing and developed countries.

My favourite chapter is close to the start of the book and describes the physics and the mathematical models used to describe the operation of solar cells. It is clear and well arranged and is a pleasure to read. It is followed by another chapter detailing the practical aspects of designing solar cells.

I thought two additional issues deserved a place in this book. First, it would have been good to discuss the wrong but oft-repeated slight against solar cells that they do not recover over their lifetimes the energy used to make them. Second, a more esoteric, but still interesting, topic is the theoretical limit to photovoltaic efficiency. Thermodynamics limits photovoltaic efficiency to about 87% but the best achieved so far is 33%. The book includes an interesting chapter on thermo-photovoltaics, although this is not really a solar-energy technology. However, I am pleased to see it included because the fields are so closely related.

Clean Electricity from Photovoltaics is an excellent resource for its intended readership of students, scientists and technologists working in the area. It seems particularly useful for specialists in one area of photo-



A changing view of the Universe

Improvements in telescopes have been mirrored by advances in our understanding of the Universe. *Beyond Earth: Mapping the Universe* (National Geographic, \$40) is a collection of essays edited by David DeVorkin that looks

back at some 5,000 years of cosmology — including the remarkable discovery by Edwin Hubble (shown above, seated at the Hooker telescope at the Mount Wilson Observatory in California) that the Universe is expanding.

voltaics research to gain an overview of closely related ones. Importantly, it is well indexed, and includes a handy list of useful web and library references. At the very least, the book deserves a place in the library of every research institution and company working on renewable energy. I expect to be referring to my copy frequently. But it is not really a book for the general reader, who might be better off with a less technical and more introductory approach, such as that taken by Kenneth Zweibel's *Harnessing Solar Power* (Plenum, 1990). ■

Richard Corkish is in the Centre for Third Generation Photovoltaics, University of New South Wales, Sydney, New South Wales 2052, Australia.

Last words of a medical historian

Madness: A Brief History

by Roy Porter

Oxford University Press: 2002. 192 pp.

£11.99, \$22

Andrew Scull

The sudden and unexpected death of Roy Porter in March robbed the English-speaking world of one of its most prolific, colourful and talented social historians and historians of medicine. In the course of little more than a quarter of a century, Porter produced a staggering amount of scholarship in a dizzying array of fields.

From the 1980s onwards, following his move from the University of Cambridge, UK, to what was then called the Wellcome Institute for the History of Medicine in London, he devoted a considerable portion of his apparently limitless energy to the history of medicine and the history of psychiatry. Early on he published primarily on topics relating to the period in which he was most at home, the eighteenth century, including a ground-breaking reinterpretation of madness in Enlightenment England (*Mind-forg'd Manacles*, Athlone, 1987). But in the last decade of his life he ranged far more broadly, producing, for example, a history of medicine from the Stone Age to the present that encompassed not just the dominant Western tradition, but also Arabic, Chinese and Indian medicine (*The Greatest Benefit to Mankind*, HarperCollins, 1997).

Madness is probably the last book Porter wrote — I say probably, for he wrote almost as fast as most scholars read, and it is possible that there is another book still in production. This slim little volume displays several of his virtues as a historian: his wide reading, his prodigious memory, his extraordinary capacity for synthesis, his eye for an anecdote, and the sheer fun he took in telling a story and constructing a narrative. Porter was

in many ways a populist, eager to reach a wide audience, and capable of attracting many readers with a witty, graceful and accessible prose style, albeit one that at times was over-addicted to alliteration, puns and wordplay. But he was also capable of serious original research, and was not afraid to advance new and thought-provoking reinterpretations of his subjects.

Readers will find little of that serious side in *Madness*. This is in large part a reflection of his goal in writing the book: to summarize, in a very brief compass indeed, what the historian can say about “who has been identified as mad? What has been thought to cause their condition? And, what action has been taken to cure or secure them?” Such an abbreviated list of questions obviously leaves many vital and fascinating issues wholly unaddressed: the place of madness in high and popular culture; the impact of mental illness on both families and the larger community; non-Western ideas about, and responses to, the insane; and the social functions of madness and psychiatry.

Even so, Porter's chosen remit is an enormous one, and it has to be said that he has been only partially successful, even on his own terms. ‘Western’ all too often turns out to mean the English experience writ large. He does make sporadic gestures towards an international and comparative focus. There are, for example, a few pages on the physician Philippe Pinel and the French Revolution, and the founder of modern neurology, Jean-Martin Charcot, and his hysterical circus; a stab at characterizing nineteenth-century German academic psychiatry and its reductionist insistence (based on little more than faith) that madness was brain disease; a nod at America's twentieth-century embrace of a bastardized Freudianism. But again and again, both the text and the handful of illustrations return to the ground Porter



Roy Porter had a sense of fun and an accessible writing style that attracted a wide audience.

found most familiar: what the English have thought and done about the crazed from the Middle Ages to the present.

For the specialist, then, *Madness* will not be a book to savour, and for the general reader it represents only a very partial brief history, in more than one sense of the term. The book is, for the most part, a good read, and is easily digested in a single sitting. Perhaps its most useful contribution, however, may be to whet the appetite for more substantial fare, and here Porter's extended and thoughtful list of suggestions for further reading may prove to be this book's best feature. ■

Andrew Scull is in the Department of Sociology, University of California at San Diego, La Jolla, California 92093-0533, USA.

New in paperback

Rocks of Ages: Science and Religion in the Fullness of Life

by Stephen Jay Gould

Ballantine, \$12.95

The Invisible Enemy: A Natural History of Viruses

by Dorothy Crawford

Oxford University Press, £8.99

“Dorothy Crawford's book portrays viruses as unseen enemies... and this approach leads to a fascinating story of their natural history.” Albert D. M. E. Osterhaus, *Nature* **409**, 19–20 (2001)

Aeons: The Search for the Beginning of Time

by Martin Gorst

Fourth Estate, £7.99

Heisenberg and the Nazi Atomic Bomb Project: A Study in German Culture

by Paul Lawrence Rose

University of California Press, \$19.95, £13.95

“Rose's book is essentially a reinterpretation of existing literature. There is little new primary material... Despite Rose's attempts to penetrate Heisenberg's “German mentality”, the author's prosecutorial analysis gives the reader little understanding of Heisenberg as a human being, or of how difficult it was to live and work under such a regime.” Mark Walker, *Nature* **396**, 427–428 (1998)

The Science of Marijuana

by Leslie L. Iversen

Oxford University Press, £15.95

Amazing algorithms

Mark J. Schnitzer

When can a physical system be said to perform a computation? In the broadest sense, every physical system performs a computation by realizing a solution to the dynamic equations that govern its physical behaviour. However, physical computing is more interesting in the narrower domain in which a set of physical variables represents the values of another set of mathematical ones. A physical variable might be a voltage within a digital computer, the height of a stack of poker chips, or a chemical concentration in a biological cell. These variables are then dynamically transformed by physical processes, in a way that represents algorithmic manipulation of the mathematical variables.

For physical computing systems that are also biological, it is often fruitful to view the organism as the 'user' of the computation. Natural selection has created many species in which individual survival rests on effective, often remarkable, computations performed by the organism's own physiology. For example, many bats locate insect prey by emitting ultrasonic pulses and detecting the echoes. By detecting small Doppler shifts in the frequency of the returning pulses, the bat's nervous system can discern acoustic 'texture' and so distinguish prey from inanimate objects. Such elegant calculations suggest some of the selection pressures that may have

shaped the underlying computational mechanisms, and can help to guide future research.

Further understanding of such computations would come from a knowledge of how acoustic variables are represented in the bat brain, of the mathematical algorithms that describe the transformation of these variables, of the biophysical mechanisms by which the transformations are performed, and of the computational errors that these mechanisms introduce in the presence of noise. These issues also concern natural selection. Algorithms should facilitate competent, or even near-optimal, computation. Theoretical engineering tools can help us to appreciate constraints on biological designs and to evaluate computational performance in relation to physical limits. Furthermore, because computations are carried out using biological molecules and cells, the physical properties of this hardware must also have constrained natural selection.

To date, most research on biological computation has been concerned with representation. For instance, *in vivo* studies have shown how acoustic variables are represented in areas of the bat brain by the electrical 'spikes' emitted by neurons. The electrical waveforms of these spikes all have a similar shape, so the acoustic variables are probably represented by the number of spikes and their arrival times. How the arrival times of spikes can represent physical variables is an area of more general debate. Nevertheless, there is widespread agreement over one of the greatest triumphs in this arena of research: Hodgkin and Huxley's classical mathematical description of the ion-channel mechanisms that underpin spike generation.

The Hodgkin-Huxley equation can be applied to a wide range of organisms, with only slight modifications. Are other biophysical mechanisms also widespread? Although many *in vitro* mechanistic studies have examined the properties of single neurons, the biophysical properties of neuronal circuits that underlie animal behaviours such as bat echolocation are less well understood. There is tension between two common views of how computations and associated animal behaviours should best be studied.

The 'mechanistic view' is that complex computations are implemented as hierarchical combinations of simpler ones, so an understanding of basic neural mechanisms will be a key that helps to unlock many complex phenomena.

In the 'algorithmic view', complex algorithms cannot be deduced from simple mechanisms as there are emergent computational principles that cannot be found by combining

Biological computation

Natural selection has created many species in which individual survival rests on computations performed by the organism's own physiology.

biophysical components. By analogy, the mathematics involved in rendering three-dimensional graphics on a computer do not follow from the workings of transistors. Hence, we should study the representations and algorithms used by systems that exhibit rich computational behaviour. One might try to deduce the algorithms used by the brain to recognize objects in a visual scene by manipulation of visual stimuli and analysis of neuronal spike patterns.

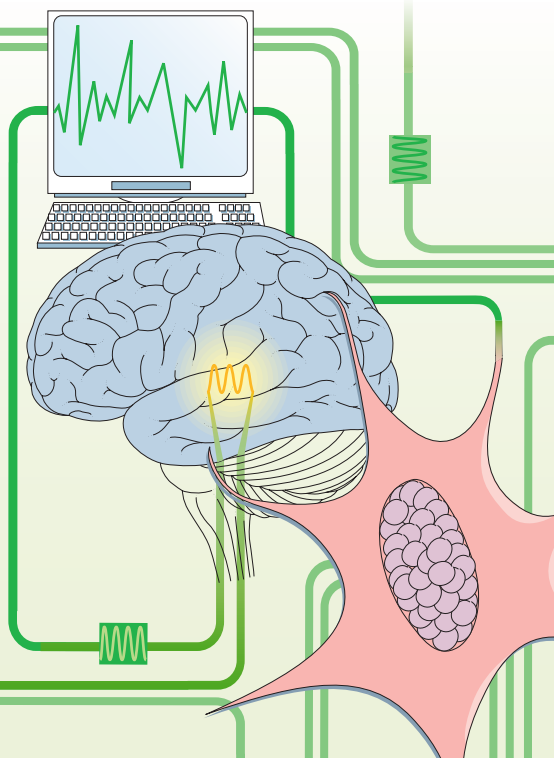
There is merit in both views, and no doubt a variety of research strategies will be needed to understand the range of biological computations. Yet it is interesting to examine the history of other areas of biological research, in which debates over the relevance of simple models to complex phenomena are more mature. What has been remarkable, and largely unanticipated, is the success of mechanistic discoveries, often concerning molecular genetics, in unravelling intricate phenomena and unifying seemingly distant research areas. Examples include the importance of yeast genetic studies for human disease and of fruitfly embryonic studies for human development.

A shared evolutionary history has often given rise to similar molecular and genetic mechanisms across a wide range of organisms, allowing us to study experimentally amenable systems and to apply the findings to other organisms. However, the physical properties that underpin the algorithms used in the brain are probably established through a complex mixture of genetic and environmental influences, so the degree to which common mechanisms will prevail remains to be seen. ■

Mark J. Schnitzer is in the Department of Biological Computation and Theoretical Physics, Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA.

FURTHER READING

Carew, T. *Behavioral Neurobiology* (Sinauer, Sunderland, Massachusetts, 2000).
von Neumann, J. *The Computer and the Brain* 2nd edn (Yale Univ. Press, New Haven, Connecticut, 1958).
Koch, C. *Biophysics of Computation* (Oxford Univ. Press, Oxford, 1999).
Weiner, N. *Cybernetics* 2nd edn (MIT Press, Cambridge, Massachusetts, 1961).
Wandell, B. A. *Foundations of Vision* (Sinauer, Sunderland, Massachusetts, 1995).



Defective promise in photonics

T. Andrew Taton and David J. Norris

Using self-assembly to produce technologically useful photonic devices becomes more feasible with the demonstration of defect engineering in a self-assembled photonic crystal.

Just as research in semiconductors led to a revolution in electronics, new materials known as photonic crystals promise a similar revolution in photonics — where photons of light rather than electrons are the fundamental carriers of information. Photons are already preferred to electrons in modern computer and communication networks because of the efficiency of sending optical signals along optical fibres. In the future, miniaturized photonic ‘chips’, or optical integrated circuits, will be needed to manipulate and process these signals. Integrated optoelectronic microchips might then replace the bulky and expensive devices that are currently used to interconvert optical and electrical information.

The chips could be made from photonic crystals, the photonic equivalent of a traditional semiconductor. In particular, these crystals exhibit a photonic bandgap^{1,2} that determines which photon wavelengths can propagate through the crystal (analogous to the electronic bandgap that governs all semiconductor devices). By introducing specific defects into this bandgap structure, the flow of photons inside the crystal can be manipulated and a variety of micrometre-scale optical devices could be fabricated³.

Self-assembly methods have been explored as a simple and inexpensive route to make photonic crystals. Unfortunately, until now, no one has been able to combine the self-assembly approach with defect engineering — a critical requirement for real applications. Writing in *Advanced Materials*, Lee *et al.*⁴ suggest a solution. They use optical microscopy to pattern imperfections in a self-assembled photonic crystal. Because their method retains the simplicity of self-assembly, it suggests a possible approach to mass-fabricating miniaturized photonic devices.

To make a photonic bandgap crystal, the first step is to place micrometre-size ‘atoms’ periodically on a three-dimensional lattice. One way to achieve this is through self-assembly, using a technique that emulates the natural process by which gemstone opals are formed. Using various laboratory tricks^{5–8}, microspheres (for example, silica colloids) are induced to organize spontaneously into a close-packed crystalline array, referred to as a ‘synthetic opal’. Next, another material (such as silicon) is deposited in the array, filling the gaps between the microspheres. Then, etching away the

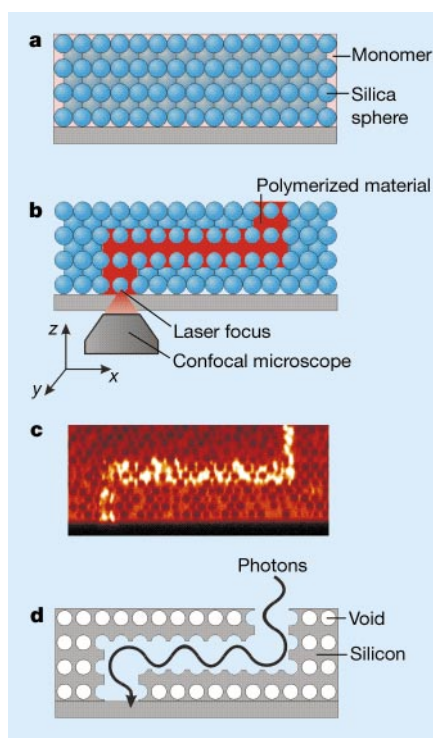


Figure 1 Making light work. Photonic bandgap crystals, formed by self-assembly, could be the basis of future photonic devices. But to build a range of workable devices, controlled defects must be engineered in the structures. By introducing a polymerization stage into the crystal-building process, Lee *et al.*⁴ have found a possible solution. **a**, After forming an ordered array of micrometre-scale silica spheres, the remaining space is filled with a photosensitive monomeric material. **b**, A laser beam focused through a confocal optical microscope polymerizes small volumes of the material, effectively ‘writing’ a defect line inside the structure in three dimensions. **c**, After unexposed monomer has been removed, staining the structure with a fluorescent dye shows up the engineered defect (image courtesy of P. Braun). **d**, Filled with a material such as silicon, and with spheres and polymer selectively removed, the resulting self-assembled photonic bandgap crystal forms a simple component that controls the passage of photons.

microspheres leaves a periodic structure that forms a photonic bandgap crystal.

But making perfect crystalline arrays of spheres is not easy. Invariably, random (bad) defects occur within the lattice. Much effort has focused on eliminating these unwanted defects from the self-assembled material, and extremely ordered photonic bandgap crystals have been made^{7,8}. To use self-assembled crystals in technological applications such as optical integrated circuits, however, an efficient method of adding intentional (good) defects within the crystal is needed; just as impurity atoms control the behaviour of conventional semiconductors, structural defects can dictate the properties of photonic bandgap materials. If properly designed, defects can trap light at given locations within the crystal. This simple effect can be exploited to make many of the components necessary for optical circuits, including optical ‘wires’ and switches.

To build specific defects into a photonic crystal, Lee *et al.*⁴ began by modifying the synthetic opal after assembly. First they filled the empty space in the opal with a monomeric material that could be polymerized by the action of multiple incident photons (Fig. 1a). Then they used a confocal micro-

scope to focus a laser beam at precise positions inside the structure. Because multiphoton polymerization depends exponentially on the intensity of the laser light, polymerization only occurs within a very small volume at the focal point of the microscope. As the laser focus was moved, three-dimensional defects of any desired shape could be ‘written’ inside the template (Fig. 1b). After polymerization, Lee *et al.* dissolved away the unexposed monomer to leave behind photogenerated patterns within the opal (Fig. 1c).

In fact, multiphoton polymerization has already been used to form exquisitely complex three-dimensional structures⁹, including a fanciful ‘micro-bull’ with horns one micrometre long¹⁰. The particular advantage of Lee *et al.*’s approach is to avoid using photopolymerization as the method for fabricating the entire photonic structure. Rather, it is used only to make small modifications to a structure that was already formed by self-assembly.

Before this approach to defect engineering can be applied in photonic bandgap structures, the feasibility of a number of steps still needs to be demonstrated. Once the synthetic opal has been modified by

polymerization, it must then be filled with a material that can support a photonic bandgap. However, many common filling techniques (such as chemical vapour deposition of silicon^{7,8}) must be carried out at temperatures that are too high for the polymerized design to survive. Various groups are already exploring ways of reinforcing similar polymeric microstructures¹¹, so this problem might soon be solved. If so, after the structure is filled with silicon (or other similar material), the synthetic opal could then be selectively removed to produce a photonic bandgap crystal with an engineered defect. For example, hollow line defects (Fig. 1d) could act as the optical wires in a future photonic chip.

Although some engineering issues remain, these results imply that, by adding a simple photopolymerization step, one of the most serious challenges to the application of self-assembled photonic crystals may soon be overcome. Lee *et al.*'s approach could have a technological impact on various applications in photonics, including miniaturized telecommunications and Internet devices. But the first effect could be to simplify basic photonics research: a researcher could design a photonic circuit, make an opal

template by self-assembly, write the circuit using the multiphoton polymerization method and test the resulting device's optical properties — all in a single day. Like silicon lithography in electronics, defect engineering in photonic crystals suggests a bright future for photonics, in which almost everything falls naturally into place. ■

T. Andrew Taton is in the Department of Chemistry and David J. Norris is in the Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455, USA.
e-mails: taton@chem.umn.edu
dnorris@umn.edu

1. Yablonovitch, E. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
2. John, S. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
3. Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. *Nature* **386**, 143–149 (1997).
4. Lee, W., Pruzinsky, S. A. & Braun, P. V. *Adv. Mater.* **14**, 271–274 (2002).
5. Holland, B. T., Blanford, C. F. & Stein, A. *Science* **281**, 538–540 (1998).
6. Wijnhoven, J. E. G. J. & Vos, W. L. *Science* **281**, 802–804 (1998).
7. Blanco, A. *et al. Nature* **405**, 437–440 (2000).
8. Vlasov, Yu. A., Bo, X.-Z., Sturm, J. C. & Norris, D. J. *Nature* **414**, 289–293 (2001).
9. Cumpston, B. H. *et al. Nature* **398**, 51–54 (1999).
10. Kawata, S., Sun, H.-B., Tanaka, T. & Takada, K. *Nature* **412**, 697–698 (2001).
11. Campbell, M., Sharp, D. N., Harrison, M. T., Denning, R. G. & Turberfield, A. J. *Nature* **404**, 53–56 (2000).

the expression of specific genes, including the gene encoding c-Fos⁵. RANKL is produced by and resides on the surface of osteoblasts. So, when an osteoclast precursor encounters an osteoblast, the resulting interaction between RANK and RANKL stimulates the osteoclast precursor to mature into a fully differentiated, bone-resorbing osteoclast.

Osteoblasts also produce osteoprotegerin, a soluble 'decoy' receptor that binds to RANKL and prevents it from binding to RANK. This effectively inhibits RANKL-mediated osteoclast maturation. Many signals that regulate osteoclast maturation do so indirectly by controlling the production of RANKL or osteoprotegerin by osteoblasts. The balance between the osteoclast-promoting RANKL and osteoclast-inhibiting osteoprotegerin can therefore regulate the number and activity of osteoclasts.

It now appears, however, that osteoclasts also have a say in their own destiny. Takayanagi *et al.*² propose a feedback mechanism by which osteoclasts control their own differentiation and function after being stimulated by RANKL (Fig. 1). Surprisingly, this signalling pathway involves the osteoclasts expressing and secreting interferon- β — a protein previously known mainly for its role in cellular responses to viruses. The proposed pathway is well supported by the authors' studies of mice (or cells derived from mice) that had been genetically engineered to lack one or other of the genes encoding various components of the pathway. So even the critical reader should be left in little doubt that this mechanism is at work *in vivo*.

Specifically, Takayanagi *et al.* propose that the activation of RANK by RANKL induces the expression of interferon- β in osteoclast precursor cells via the transcription factor c-Fos. Interferon- β is then released from these cells and binds to and activates its own cell-surface receptor system on osteoclast precursors, leading to a decrease in c-Fos levels (Fig. 1). Without enough c-Fos, osteoclast differentiation is inhibited, bringing this negative-feedback mechanism full circle.

Several key observations serve as the basis for this model. First, Takayanagi *et al.* found that mice lacking the interferon- β receptor have low bone mass and show increased bone resorption by osteoclasts, implying that the osteoclasts are not being controlled appropriately. Mice that lack interferon- β have a similar defect. These findings suggest that interferon- β inhibits the differentiation or function of osteoclasts.

Next, Takayanagi *et al.* showed that RANKL induces the expression of interferon- β in cultured osteoclast precursor cells. The authors also provide evidence that c-Fos can directly activate expression of the interferon- β gene. Furthermore, adding

Medicine

Interfering with bone remodelling

Tamara Alliston and Rik Derynck

As they mature, bone-resorbing cells trigger the production of their own 'off-switch' — the interferon- β protein — to prevent the runaway bone loss that is seen in diseases such as osteoporosis.

In adult vertebrates, ten per cent of the skeletal bone mass is replaced every year, amounting to a complete structural overhaul every decade. This constant remodelling allows bone to carry out its many functions: to support the body and allow movement; to incubate developing immune cells; and to act as a reserve of inorganic minerals, especially calcium. Remodelling repairs bone defects that result from the stress of constant use, and helps to maintain the optimal levels of calcium in the blood that are required for cells to function.

The contractors in this remodelling project include cells that destroy and resorb old bone (osteoclasts), and those that deposit new bone in its place (osteoblasts). Their activities must be carefully controlled, because a slight imbalance between bone destruction and formation can have dire consequences, as is seen in osteoporosis. Several molecules have been proposed to coordinate the function of osteoclasts with that of osteoblasts¹, but osteoclasts may also keep themselves under strict control. On page 744 of this issue, Takayanagi and colleagues²

show that these cells can do so by drawing upon a molecule — interferon- β — that is traditionally produced by immune cells in response to viruses. The authors further suggest that interferon- β might be useful for slowing the excessive bone destruction that occurs in some human diseases.

To better understand the details of these results², readers will need to know a little about the mechanisms that have already been found to control osteoclast maturation (differentiation). One of the molecules required is the transcription factor c-Fos, which can bind DNA and drive the expression of genes needed for osteoclasts to function³. Accordingly, mice that lack a functional c-Fos gene develop osteopetrosis — a disease characterized by decreased bone resorption and increased levels of calcified bone matrix.

Two cell-surface proteins, RANK and its partner RANKL (for 'RANK ligand'), are also key regulators of osteoclast formation and function⁴. RANK is present on osteoclast precursor cells and, when activated, promotes osteoclast maturation by increasing

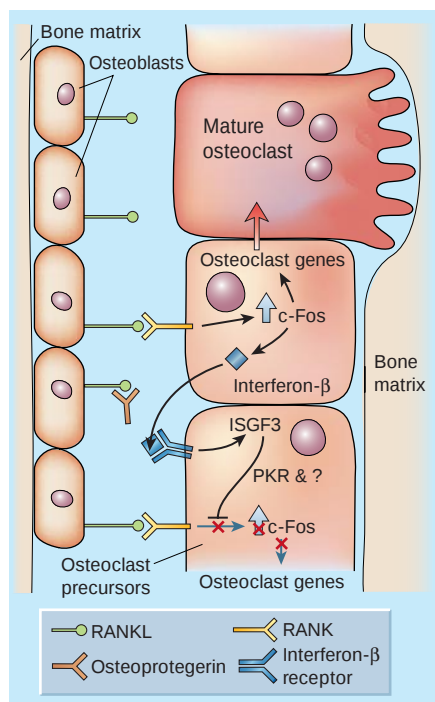


Figure 1 Balancing bone remodelling. The activity of osteoclasts — cells that destroy and resorb bone — is kept in check in several ways. Osteoblasts (cells that make new bone, left) produce the RANKL protein, which activates RANK on osteoclast precursor cells, stimulating these cells to differentiate into mature, bone-resorbing osteoclasts. Activated RANK does this by inducing the expression of c-Fos, which binds to DNA and activates expression of several genes needed for osteoclast function. Takayanagi *et al.*² find that c-Fos also activates the expression of the interferon- β gene. The interferon- β protein is then secreted and binds its receptor, presumably on neighbouring osteoclast precursors, inhibiting osteoclast differentiation by preventing the RANK-induced expression of c-Fos. ISGF3 and PKR are among the proteins that are intermediaries between the interferon- β receptor and c-Fos. Osteoprotegerin is a soluble protein, released from osteoblasts, that binds to RANKL and prevents it from binding to RANK.

interferon- β to cultured osteoclast precursors prevents the RANKL-induced increase in c-Fos protein levels without affecting other responses to RANKL, and inhibits the RANKL-induced differentiation of the cells into osteoclasts. This inhibition can be overcome by experimentally forcing increased c-Fos expression. Finally, the authors reveal that interferon- β reduces osteoclast-mediated bone destruction in a mouse model.

Takayanagi *et al.*'s observations have profound implications for our understanding of bone remodelling. They also reveal a new role and a new physiological context for interferon- β , which has not previously been linked to osteoclast differentiation⁶. And finally, the authors have uncovered a new mechanism by which the expression of interferon- β is activated — c-Fos was not known to be involved in such activation, and indeed interferon- β expression in response to viral infection is induced by other transcription-factor complexes.

In retrospect, we should perhaps not be too surprised to learn that interferon- β regulates osteoclast differentiation. After all, bone is the site of maturation of certain types of immune cells, and osteoclasts are derived from the same progenitor cells as some immune cells. Moreover, RANKL-related signalling pathways are involved in several immune responses, and RANKL itself is expressed by the T cells of the immune system⁷. In addition, interferon- γ , which is distantly related to interferon- β and signals through a different receptor, also inhibits RANKL-induced signalling and osteoclast maturation⁸, albeit through a different mechanism (involving degradation of a signalling protein downstream of RANK). As

interferon- γ is secreted by T cells but not by osteoclasts, its osteoclast-inhibitory activity is not part of a negative-feedback mechanism, but rather illustrates regulatory crosstalk between the immune system and bone remodelling. These and other studies reveal the need for more research into the interplay between immune cells (and their regulators) and bone.

Evolutionary biology

The tale of the parasitic cuckoos

Arie J. van Noordwijk

These days, investigations of evolutionary events in groups of organisms can be taken well beyond the just-so story. Analysis of how members of the cuckoo family became 'brood parasites' provides a wonderful example.

Birds of the cuckoo family show a fascinating variety in parental care. There are 53 species, called brood parasites, that lay their eggs in the nests of other bird species; the 83 other cuckoo species raise their young themselves, although some of them may also try to lay eggs in the nests of birds of their own species. The outcome of parasitism is that the offspring of the unfortunate hosts die in one way or another, cuckoo young being raised instead (Fig. 1, overleaf).

So parasitic cuckoos and their hosts have a conflict of interest, and throughout evolutionary history they will have been vying for advantage in a coevolutionary arms race^{1,2}. For instance, whereas the tendency in parasitic cuckoos will be to make their eggs as much like their hosts' eggs as possible, so as to fool the prospective step-parents,

Finally, the new observations² suggest that the targeted use of interferon- β might be beneficial for treating diseases such as osteoporosis, rheumatoid arthritis and periodontal disease, in which bone resorption is inappropriately high⁹. Takayanagi *et al.*'s discovery that interferon- β prevents bone resorption in a mouse model of inflammatory arthritis supports this idea. Perhaps much could already be learned from a close examination of bone remodelling and metabolism in people, such as those suffering from multiple sclerosis, who are undergoing long-term treatment with interferon- β . Indeed, some such patients have shown improvement in their rheumatoid arthritis¹⁰. Clearly, Takayanagi *et al.*'s results will interest scientists and physicians alike.

Tamara Alliston and Rik Derynck are in the Departments of Growth and Development, and Anatomy, University of California at San Francisco, San Francisco, California 94143-0640, USA.

e-mails: tallist@itsa.ucsf.edu

derynck@itsa.ucsf.edu

1. Karsenty, G. *Genes Dev.* **13**, 3037–3051 (1999).
2. Takayanagi, H. *et al. Nature* **416**, 744–749 (2002).
3. Grigoriadis, A. E., Wang, Z. Q. & Wagner, E. F. *Trends Genet.* **11**, 436–441 (1995).
4. Suda, T. *et al. Endocrinol. Rev.* **20**, 345–357 (1999).
5. Matsuo, K. *et al. Nature Genet.* **24**, 184–187 (2000).
6. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. *Annu. Rev. Biochem.* **67**, 227–264 (1998).
7. Kong, Y. Y., Boyle, W. J. & Penninger, J. M. *Immunol. Today* **21**, 495–502 (2000).
8. Takayanagi, H. *et al. Nature* **408**, 600–605 (2000).
9. Rodan, G. A. & Martin, T. J. *Science* **289**, 1508–1514 (2000).
10. Tak, P. P. *et al. Rheumatology* **38**, 362–369 (1999).

many hosts will fight back by recognizing the cuckoo eggs and throwing them out of the nest.

How did this arms race begin? This is a question that has been tackled by Krüger and Davies, as they describe in *Proceedings of the Royal Society*³. Their starting point is the view that the ancestral cuckoos were non-migratory birds that inhabited tropical forests and laid big eggs which they cared for themselves. This view is based on the observation that non-parasitic members of the cuckoo family, and all their close relatives outside the family, have these characteristics. Krüger and Davies then analysed data on 13 traits related to ecology and life history for all 136 cuckoo species, to find out which of these characteristics predated the evolution of parasitism and which evolved as consequences of parasitism. Apart from



Figure 1 Feathered parasite — a young cuckoo (*Cuculus canorus*) is fed by its unwitting foster parent, a hedge sparrow (*Prunella modularis*).

migratory behaviour and egg size, the variables analysed included body weight, diet and habitat.

Such a reconstruction of evolutionary events is based on two elements. First, one needs a reliable phylogeny — a scheme of relationships between species — to provide a detailed reconstruction of which species had more recent and which more distant common ancestors. This allows evolutionary history to be summarized as a series of independent events in which single ancestral groups gave rise to two new groups (Fig. 2). The second element is a conceptual tool. Summarizing the history as a series of independent events allows us to filter out the consequences of common history⁴. At each branching point, or node, in the phylogeny, one can see whether all species in a branch have a particular trait in common, or whether there is diversity within that branch. If all members of a branch share the trait in question, the simplest explanation is that the evolutionary change that produced it took place between the node where all descendants share the trait and the previous node that combines species that subsequently separated. At each branching point where one arm consists of parasitic species and the other of non-parasitic species, we assume that there was one evolutionary change in parasitism.

These two groups can then be contrasted and compared for all other traits. When there are enough independent contrasts, statistics can be applied to the data (Fig. 2). This procedure becomes interesting when we find that some changes in life-history characteristics predate others. It is in many ways similar to uncovering a metabolic pathway by identifying the various steps through which the metabolic process occurs. Instead of biochemical reactions, however, we are dealing

here with evolutionary transitions, which are under the influence of either natural selection or chance. The judgement that natural selection, rather than chance, has been involved can be based on the number of independent cases that have the same series of transitions.

Krüger and Davies used two cuckoo phylogenies produced by different authors. One of them is based on bone characteristics⁵ of cuckoo species and suggests that brood parasitism among birds of different species evolved just once. The other is based on molecular traits⁶ and implies that brood parasitism had three independent origins. These numbers are clearly too small to allow

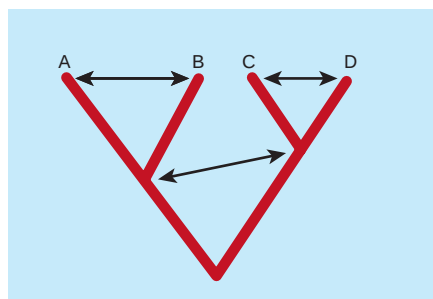


Figure 2 Independent contrasts of traits in evolutionary biology. Given a phylogenetic tree relating species, the central idea is that comparisons between species A and B, and between C and D, or comparison of the average of A and B with the average of C and D, can reveal independent evolutionary events. In the case of Krüger and Davies' study³, for example, if species A and B are parasitic and C and D are not, only one evolutionary change has taken place. If A and C are parasitic, and B and D are not, there have been two evolutionary changes. If there are sufficient independent events, statistical analysis can be applied to draw conclusions about trait evolution.

a statistical analysis. Importantly, however, there are intermediate steps in brood parasitism — between cuckoo species that cannot raise their own young (obligate parasites) and species that always raise their own young, there is a group of cuckoos that can raise their young but will try to lay their eggs in the nests of birds of the same species (facultative parasites). Using the intermediate steps, Krüger and Davies found that 34 or 19 contrasts could be made between facultative and non-parasitic groups for the phylogenies based on bone or molecular traits, respectively. There are some differences in the details of the results depending on the phylogeny used, but the broad picture is the same.

Overall, Krüger and Davies can say that increased degrees of brood parasitism go together with smaller body size, smaller eggs, eating smaller prey, a more open and less productive habitat, a larger breeding area and a more migratory lifestyle. As an example of a biological explanation for these changes, the authors point out that brood parasitism means that a cuckoo will expend less energy because it does not have to feed its own young. So such species can afford to inhabit a less productive habitat, with lower food availability, than they would have otherwise done.

But Krüger and Davies go further, and conclude that changes in diet and range area, and becoming migratory, preceded the evolution of brood parasitism, whereas changes in egg size occurred later and were a consequence of host-parasite coevolution. Two decades ago, such a statement could have been criticized as a simple just-so story, stemming from an adaptationist agenda intent on regarding all biological characteristics as the result of adaptations to environment. Nowadays it can be seen to have rigorous statistical support.

It is, of course, nice to be more confident that an evolutionary story is likely to be true. The real significance, however, is that this procedure — attaching probabilities to each transitional step in potential evolutionary sequences — helps us to translate observed patterns of change into chains of causal processes. This in turn allows us to make predictions and to bring together different situations where the processes involved are similar or the same. For instance, brood parasitism occurs not only in birds but in several insects (one example being the 'cuckoo bee'). It is virtually impossible to compare the phenomenon in different types of organism by comparing observed patterns. But it is likely that the similarities among the causal processes are much more obvious, and will allow us to see common denominators.

There is a third way in which Krüger and Davies' study³ should be influential. It shows that we can probably learn much

more from studying intermediate stages of brood parasitism than from investigating the advanced cases where coevolution has already mutually adapted the parasite's and the host's life history and behaviour. Obligate parasitic cuckoos have long been a favourite of students of evolution and behavioural ecology, and I doubt that the flow of papers concerning these birds will slacken. However, we may well see an increase in studies of the intermediate — facultative — brood parasites, and they

should be especially illuminating.

Arie J. van Noordwijk is at the Netherlands Institute of Ecology, PO Box 40, NL6666ZG Heteren, The Netherlands.

e-mail: noordwijk@ccto.nioo.knaw.nl

1. Takasu, F. *Evol. Ecol.* **12**, 969–987 (1998).
2. Davies, N. B., Bourke, A. F. G. & Brooke, M. D. *Trends Ecol. Evol.* **4**, 274–278 (1989).
3. Krüger, O. & Davies, N. B. *Proc. R. Soc. Lond. B* **269**, 375–381 (2002).
4. Pagel, M. *Nature* **401**, 877–884 (1999).
5. Hughes, J. M. *Zool. J. Linn. Soc.* **130**, 263–307 (2000).
6. Aragon, S. *et al. J. Evol. Biol.* **12**, 495–506 (1999).

Climate change

The 20-year forecast

Francis W. Zwiers

Policy-makers need short-term climate predictions to develop strategies for coping with climate change over the typical two-decade planning horizon. Two new studies increase our confidence in these predictions.

Earth's climate has changed during the past century^{1,2}. The global mean temperature has risen by 0.6 ± 0.2 K; there have been reductions in snow cover, glacier mass and sea-ice extent; and sea level and the heat held by the oceans have increased. This observational evidence, and sophisticated detection studies that have identified the global signature of an anthropogenic effect in the twentieth-century record, have led the Intergovernmental Panel on Climate Change (IPCC) to conclude¹ that “most of the warming observed over the past 50 years is attributable to human activities”. All projections of future change indicate that the warming is likely to continue. This conclusion holds regardless of the computer model used or the ‘emission scenario’ — the particular set of data describing the future emissions of greenhouse gases and aerosols — applied in the model.

Climate projections have often looked to the year 2100. This is scientifically valuable because it highlights differences between projections obtained using different models. It may also reveal gaps in our understanding of processes that critically affect how the climate responds to external influences such as the anthropogenic emission of greenhouse gases. However, 2100 is well beyond the typical two- to three-decade horizon for developing policy, and for planning and implementing strategies to mitigate or adapt to climate change. The attention paid to estimates for 2100 has also obscured the message that there is considerable agreement on projections for the next two to three decades — and that the agreement is independent of the particular model or emission scenario applied, especially when the projections take into account changes observed in the past century.

New papers by Knutti *et al.*³ and Stott and

Kettleborough⁴ (pages 719 and 723 of this issue) consider climate change throughout the twenty-first century. But their particular value is to add considerably to the developing consensus on the projected global mean change two to three decades from now. Moreover, like several other recent analyses^{1,5,6}, uncertainty in the projections is presented in probabilistic terms by including estimates of the likely range of future warming that are consistent with observed twentieth-century change. Putting information in this format is helpful for planning and policy development: users can begin to assess the expected losses, costs and benefits that might

occur in the absence or presence of measures to mitigate or adapt to change over a range of likely outcomes.

Knutti *et al.* and Stott and Kettleborough take very different approaches. Stott and Kettleborough⁴ use a comprehensive, coupled atmosphere–ocean global climate model. This type of model simulates the circulation of the atmosphere and ocean as a whole. It produces natural variability in weather and climate, on timescales from hours to centuries, similar to that observed; and it incorporates the main feedback mechanisms that are thought to have determined the climatic response to natural influences (such as variations in solar output and the occurrence of explosive volcanic activity), as well as to anthropogenic greenhouse-gas and aerosol emissions during the twentieth century¹. These models are expensive to run, and so only comparatively few simulations of past and future climate can be performed.

Stott and Kettleborough apply the regression methodology used in studies^{7,8} to detect climate change, and attribute cause to that change. This allows them to scale up signals estimated from small ensembles of simulations of twentieth-century conditions so that they best match the observed historical change. The scaling factors are then used to adjust the model's projections of future change. An uncertainty range is estimated by accounting for uncertainty in the scaling factors and the effects of natural variability.

Stott and Kettleborough estimate that the global mean temperature in the decade 2020–30 will be 0.3–1.3 K greater than in 1990–2000 (5–95% likelihood range). This

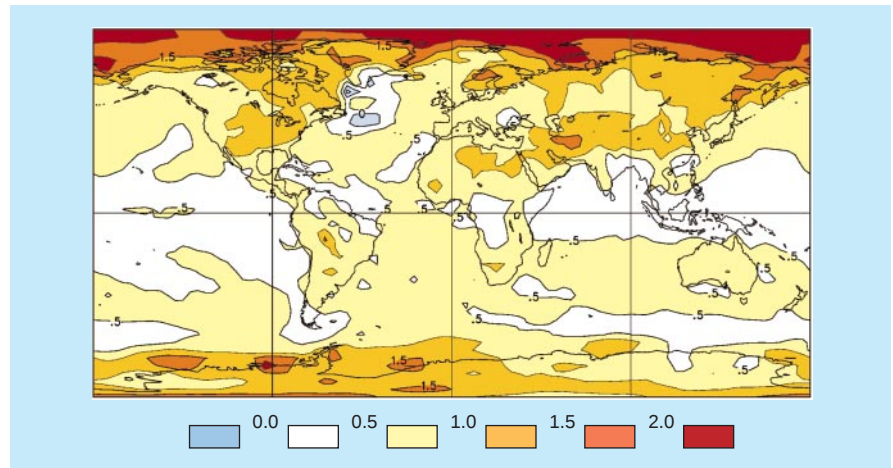


Figure 1 A warmer world. Shown here are projected changes in surface air temperature, relative to 1990–2000, for the decade 2020–30. The global change in mean temperature estimated by this model¹⁵, the Canadian Centre for Climate Modelling and Analysis CGCM2, is 0.68 K. This result is similar to that obtained by other atmosphere–ocean global climate models, such as that used by Stott and Kettleborough⁴. It is also similar to the conclusions drawn by Knutti *et al.*³ from a model of intermediate complexity. The change projected by CGCM2 falls well within the approximate 5–95% uncertainty ranges estimated by Stott and Kettleborough (0.3–1.3 K) and by Knutti *et al.* (0.5–1.1 K). The global change in mean temperature is expected to have an uneven distribution geographically, with generally greater warming over land and at high latitudes than elsewhere. So local effects of climate change may be greater than one might infer from global mean projections.

result is unaffected by the choice of emission scenario used to make the projection. Indeed, temperature projections produced using high- and low-emission scenarios do not diverge significantly until mid-century. That is partly because the large thermal inertia of the oceans means that a good proportion of the change in the next two decades will result from the climate's adjustment to changes in the greenhouse-gas content of the atmosphere that have already occurred. It is also partly because the effects of greenhouse-gas and sulphur-dioxide emissions offset each other in a similar way in the various scenarios. Sulphate aerosols produced from sulphur-dioxide emissions partially counteract the warming effect of greenhouse-gas emissions — both directly, by reflecting some solar radiation back into space, and indirectly, by increasing the reflectivity and longevity of clouds.

Knutti *et al.*³ use a climate model of intermediate complexity⁹. It has an atmospheric component that represents north–south variations in the distribution of heat and moisture, and an oceanic component that accounts for north–south and vertical motion of water, and variations in salinity and heat content. Because of its simplicity, this model, and others like it, runs very quickly, allowing a thorough analysis of variations in the model's parameters.

Knutti *et al.* pursue this approach on a grand scale. Beginning with a broad distribution of settings that are consistent with previous studies using the method^{5,10,11}, they screen thousands of randomly selected parameter combinations to identify settings that produce changes in surface warming and ocean heat content that are consistent with those observed in the past century. They then extend these simulations forward in time using particular emission scenarios — the SRES A2 and B1 scenarios¹². What they obtain are probability distributions of future temperature change that are consistent with past changes in surface temperature and ocean heat content.

Overall, their conclusions agree with those of Stott and Kettleborough. Knutti *et al.* find that the projected distribution of likely surface warming is independent of the choice of emission scenario for the next several decades; that the probable warming for 2020–30 relative to 1990–2000 is about 0.5–1.1 K (5–95% likelihood range); that the upper range of the warming uncertainty is probably greater than the IPCC range¹; and that 'scenario uncertainty' (differences resulting from the particular choice of emission scenario) and 'model uncertainty' (as embodied in the range of parameter settings that are consistent with observed climate change) contribute about equally to the range of uncertainty in projected change at the end of the twenty-first century.

These results, obtained using very differ-

ent methods, present a consistent picture of future climate, both within the two- to three-decade planning horizon (Fig. 1) and for the end of the twenty-first century. For the near term, projections made by different models with different emission scenarios produce remarkably similar results on the global scale. Furthermore, presenting results probabilistically gives policy-planners the information they need to assess expected costs, losses and benefits over a range of probable outcomes. In the longer term, these and other studies^{1,5} show that model and scenario uncertainty are roughly equal contributors to the total uncertainty about warming by 2100.

Neither Knutti *et al.* nor Stott and Kettleborough take into account factors such as 'structural uncertainty'⁵ (uncertainty that arises because models may not accurately represent all climatically important processes), carbon-cycle feedbacks¹³ or rapid, nonlinear climate change¹⁴. Uncertainties therefore remain that are beyond the statistical uncertainties described in the two papers. But both sets of authors point out that the upper bound on the potential warming for 2100 may well be above the IPCC figure¹ of 5.8 K under 'heavy emissions' scenarios.

DNA repair

Breaking the seal

Sebastian D. Fugmann

The diversity of the receptors on our immune cells that recognize 'foreign' material is ensured by combining a set of gene segments to form the final receptor genes. A crucial player in that process has now been found.

One hallmark of our immune system is its ability to recognize and react efficiently to a nearly infinite variety of infectious microorganisms. Such remarkable versatility stems from a cut-and-paste process that rearranges the genome of immune cells as they develop. These cells — B and T lymphocytes — detect pathogen molecules through 'antigen-receptor' proteins and, by assembling the genes encoding these receptors from different DNA segments, the cells ensure that the receptors themselves come in a huge variety of forms.

During the assembly process, a pair of lymphocyte-specific scissors accurately excises the regions between such receptor-encoding gene segments, and a general DNA-repair machinery joins the resulting 'loose' DNA ends together. But it has long been unclear how these two processes are connected. Two years ago¹ it was discovered that the gene encoding a protein named Artemis is mutated in people who, because of a lack of T and B cells, suffer severe immunodeficiency. Writing in *Cell*, Ma *et al.*² now provide convincing evidence that the

So policy-makers should not discount the possibility of a very warm climate when considering long-range policy options. ■

Francis W. Zwiers is at the Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada, c/o University of Victoria, PO Box 1700, Victoria, British Columbia V8W 2Y2, Canada.

e-mail: Francis.Zwiers@ec.gc.ca

1. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
2. Levitus, S. *et al.* *Science* **292**, 267–270 (2001).
3. Knutti, R., Stocker, T. F., Joos, F. & Plattner, G.-K. *Nature* **416**, 719–723 (2002).
4. Stott, P. A. & Kettleborough, J. A. *Nature* **416**, 723–726 (2002).
5. Wigley, T. M. L. & Raper, S. C. B. *Science* **293**, 451–454 (2001).
6. Allen, M. R., Stott, P. A., Schnur, R., Delworth, T. & Mitchell, J. F. B. *Nature* **407**, 617–620 (2000).
7. Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. & Mitchell, J. F. B. *Nature* **399**, 569–572 (1999).
8. Stott, P. A. *et al.* *Clim. Dyn.* **17**, 1–21 (2001).
9. Claussen, M. *et al.* *Clim. Dyn.* **18**, 579–586 (2002).
10. Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. & Webster, M. D. *Science* **295**, 113–117 (2002).
11. Andronova, N. & Schlesinger, M. E. *J. Geophys. Res.* **106**, 22605–22612 (2001).
12. Nakicenovic, N. & Swart, R. *Special Report on Emissions Scenarios* (Cambridge Univ. Press, 2000).
13. Cox, P. N., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. *Nature* **408**, 184–187 (2000).
14. Broecker, W. S. *Science* **278**, 1582–1588 (1997).
15. Flato, G. M. & Boer, G. J. B. *Geophys. Res. Lett.* **28**, 195–198 (2001).

Artemis protein is the catalytic subunit of an enzyme that might be the long-sought missing link between DNA cutting and pasting.

The assembly of antigen-receptor genes from individual gene segments — 'V(D)J recombination' — is one of four known processes in which a mammalian cell, in this case a developing B or T cell, actively modifies its own genome. (The three others are targeted to the recombined antigen-receptor genes in B cells, and are responsible for fine-tuning pathogen recognition and subsequent cellular responses.) V(D)J recombination is, of course, advantageous in that it ensures maximal diversity of antigen receptors. But it also poses a risk for the cell. If the process runs out of control — if the loose DNA ends are not joined, for example — it can lead to genomic aberrations and, eventually, to cancer. So the lymphocyte-specific scissors have to communicate efficiently with the ubiquitous DNA-repair machinery to ensure a tightly regulated transfer of the intermediate DNA ends. Studies in numerous labs have yielded an elaborate model for the DNA-cutting step³. But the joining phase

NATURE

100 YEARS AGO

In the year 1854 it was proved by Joule and Lord Kelvin that hydrogen on free expansion behaved differently to all other gases. Air, when allowed to expand from a higher to a lower pressure without performing external work, became cooled, the fall of temperature being proportional to the difference of pressure; hydrogen, on the other hand, became warmer. As is well known, the Joule–Kelvin effect has been applied by Hampson and Linde to the liquefaction of air in quantity, but since for hydrogen the effect is of opposite sign, it was obvious that the Hampson–Linde apparatus could not be directly applied to the liquefaction of that gas; a conclusion which is confirmed by this research. There seemed no doubt, however, that after due modification the Hampson–Linde method would be applicable to the liquefaction of hydrogen, the necessary condition being that the gas should be cooled to below the temperature at which the Joule–Kelvin effect changes sign... From these experiments it appears that the temperature of the inversion of the Joule–Kelvin effect for hydrogen lies about $-80^{\circ}5\text{ C.}$, a number which agrees very closely with that arrived at by Witkowski from the Rose–Innes equation ($-79^{\circ}3$).

From *Nature* 17 April 1902.

50 YEARS AGO

The first clearly proved case of total sex-linkage in the house mouse has been found — a fact worthy of record since very few sex-linked genes are yet known in mammals other than man, and none in the rodents. Apart from man, where hæmophilia and colour-blindness are the best known of a number of more or less certainly sex-linked genes, there are only yellow in cats and hæmophilia in dogs. In the mouse two genes, *mottled* and *brindled*, have been described which may be totally sex-linked; but this cannot be directly proved as both genes are lethal in males... The new gene is semi-dominant and behaves in every way as would be expected of a totally sex-linked gene with no homologue on the Y-chromosome, though the numbers of mice bred are not yet enough to detect minor deviations from expectation... Heterozygous females have transverse black stripes reminiscent of the tabby markings of cats, and for this reason the gene is named '*tabby*'.

From *Nature* 19 April 1952.

of $V(D)J$ recombination, and how it is mechanistically coupled to DNA cleavage, is far less well understood.

The DNA-cutting scissors consist of a complex made up of the RAG proteins, which introduce a break at each end of the non-coding DNA fragment that separates the two antigen-receptor-encoding gene segments. The ends of the excised fragment are 'blunt' and can be joined directly, producing a circular piece of DNA (Fig. 1). In contrast, each gene segment is sealed at its cut end by a covalent bond between the two strands that make up a DNA molecule. This 'hairpin' structure protects the gene segments from being degraded, but also prevents them from being joined together immediately into a functional gene.

To achieve joining, a break must first be introduced near the tip of each hairpin in one of the DNA strands (Fig. 1). The resulting DNA ends are typically not compatible and require processing before being joined. This involves a general DNA-repair machinery (known as the non-homologous-end-joining machinery) that consists of five core proteins⁴: Ku70, Ku80, DNA-PK_{cs}, Xrcc4 and DNA ligase IV. Ku70 and Ku80 bind to DNA ends as a dimer, and then associate with DNA-PK_{cs} to form an enzyme, the DNA-dependent protein kinase (DNA-PK). This complex is thought to serve as a central docking platform that recruits and regulates additional factors and facilitates the joining of DNA ends by Xrcc4 and DNA ligase IV.

But which is the enzyme responsible for opening hairpin-sealed DNA ends? The first big step towards the answer came with the

observation that the assembly of antigen-receptor genes is blocked in '*scid*' mice⁵ — severely immunodeficient animals that have a mutated DNA-PK_{cs} protein. In these mice, the gene segments remain protected by hairpin seals⁶, whereas the ends of the excised fragment are readily joined⁷. A group of humans with severe combined immunodeficiency (SCID) have a similar defect in assembling antigen-receptor genes, but surprisingly, the gene encoding DNA-PK_{cs} is unaltered in these patients⁸. Later analysis¹ revealed that the human defect results from mutation of a newly identified gene encoding the protein Artemis, the function of which remained unclear.

Ma *et al.*² have now neatly assembled the two findings into a single, simple story: the DNA-PK_{cs} and Artemis proteins form a complex that can open hairpins, as well as remove single-stranded overhangs from double-stranded DNA, in biochemical assays. Importantly, the complex was robustly active when the RAG proteins were included in the reaction to create hairpin ends *in situ*.

So, is the Artemis–DNA-PK_{cs} complex the long-sought hairpin-opening machinery? Previous biochemical studies^{9–11} showed that the RAG proteins, or a DNA-repair complex comprising the Mre11, NBS1 and Rad50 proteins, can open DNA hairpins *in vitro*. But these studies fell short of explaining why hairpin-sealed DNA ends accumulate in the absence of DNA-PK_{cs}. Ma *et al.*'s results now suggest a simple bucket-brigade model: the RAG proteins cleave the DNA and the resulting hairpin-sealed ends are transferred to the Artemis–DNA-PK_{cs}

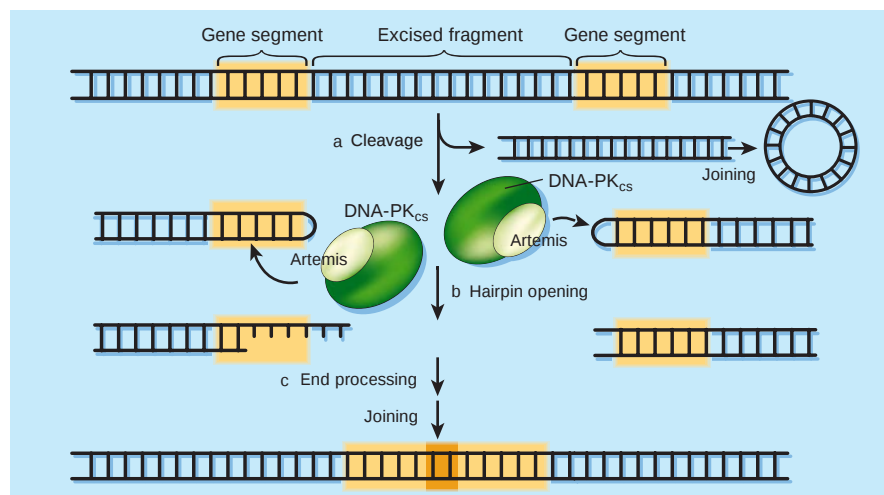


Figure 1 Genetic cutting and pasting in immune cells. This process, $V(D)J$ recombination, allows B and T cells to generate a variety of antigen receptors and thereby recognize many different pathogens. During B- and T-cell development the genes encoding antigen receptors are assembled from individual gene segments (pale yellow boxes). a, DNA is cut precisely at the border of the gene segments by the lymphocyte-specific RAG proteins. The excised DNA fragment is circularized, while the gene segments are protected by hairpin structures. b, Ma *et al.*² report that the Artemis–DNA-PK_{cs} complex (green ellipses) opens these hairpin seals at different positions near their tips. c, The ends of the gene segments are processed — presumably also in part by the Artemis–DNA-PK_{cs} complex — and finally joined. The imprecise nature and uniqueness of each resulting joint (orange box) contributes to the diversity of antigen receptors.

complex, where they are opened, processed and transferred to yet more enzymes for further processing and joining. The first tough experimental challenge for this model will be to analyse whether the ends of gene segments remain sealed by hairpin structures in Artemis-deficient cells.

Beyond *V(D)J* recombination, Artemis is also important in the general non-homologous-end-joining pathway for DNA repair. Just like cells from *scid* mice, Artemis-deficient human cells are very sensitive to chemicals or levels of irradiation that induce double-strand breaks in DNA⁸, implying that Artemis is needed to repair these breaks. But such DNA damage rarely involves hairpin-sealed DNA ends. So the predominant role of the Artemis–DNA–PK_{cs} complex here can probably be attributed to its ability to remove single-stranded overhangs from DNA ends². Incompatible overhangs, which might develop through the loss of nucleotides from initially perfectly matching DNA ends, would render these ends unable to join.

Last but not least, Ma *et al.*'s paper² brings

us closer to the ultimate goal in this field — to establish a complete *in vitro V(D)J* recombination system, using purified proteins. Such a system will allow us to study the mechanisms of DNA repair in this process, and will boost our understanding of how DNA breaks more generally are healed effectively, to maintain genomic integrity and so to prevent cancer. ■

Sebastian D. Fugmann is at the Yale School of Medicine, Section of Immunobiology, 310 Cedar Street, New Haven, Connecticut 06510, USA.
e-mail: sebastian.fugmann@yale.edu

1. Moshous, D. *et al.* *Cell* **105**, 177–186 (2001).
2. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. *Cell* **108**, 781–794 (2002).
3. Fugmann, S. D., Lee, A. I., Shockett, P. E., Vile, I. J. & Schatz, D. G. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
4. Grawunder, U. & Harbst, E. *Curr. Opin. Immunol.* **13**, 186–194 (2001).
5. Bosma, M. J. & Carroll, A. M. *Annu. Rev. Immunol.* **9**, 323–344 (1991).
6. Roth, D. B. *et al.* *Cell* **70**, 983–991 (1992).
7. Lieber, M. R. *et al.* *Cell* **55**, 7–16 (1988).
8. Nicolas, N. *et al.* *J. Exp. Med.* **188**, 627–634 (1998).
9. Besmer, E. *et al.* *Mol. Cell* **2**, 817–828 (1998).
10. Shockett, P. E. & Schatz, D. G. *Mol. Cell. Biol.* **19**, 4159–4166 (1999).
11. Paull, T. T. & Gellert, M. *Genes Dev.* **13**, 1276–1288 (1999).

Astronomy

Worlds of mutual motion

Jean-Luc Margot

Two asteroids beyond Neptune have been found to orbit one another. This binary system in the Kuiper belt shows marked differences from binary objects elsewhere in the Solar System.

The existence of asteroids between the orbits of Mars and Jupiter has been known for about 200 years. These small, irregular bodies are the remnants of planetary embryos that have been broken up by collisions to provide convenient samples of various stages in the planet-building process. About 200 of the objects in this belt are larger than 100 km in diameter, and about a million are larger than 1 km. The total mass of these asteroids is about 5×10^{-4} that of Earth.

Another set of asteroids, called the Kuiper belt, is at least a hundred times more massive and resides beyond the orbit of Neptune. Characterizing these objects is important to constrain models of planet formation, which generally involve accretion of planetesimals in a rotating disk. The sizes of the largest objects provide a handle on the timescales required to assemble them, and the total mass in today's Kuiper belt constrains the processes of mass loss since the initial accretion. Such observations also allow conditions in our Solar System to be compared with those for the disks and planets observed around other stars. These tasks have just become easier with the discovery of several binary systems in the Kuiper belt, the first of which, named 1998 WW31, is now described by Veillet *et al.* on page 711 of this issue¹.

Binary systems — two objects orbiting one another — are astronomical treasures for both the observer and the theorist. Their very existence raises perplexing questions about

their formation, stability and evolution^{2–4}. But their abundance and orbital configurations provide helpful evidence about the circumstances in which they formed, as well as the environment in which they currently reside. The characteristics of their orbit around each other, and mutual tidal influence, depend on the bulk properties of the components, and observations of their motion provide information that would otherwise only be available from spacecraft flybys. For instance, measurements of the orbital period and separation of the two bodies directly constrain the total mass of the system, according to Kepler's third law. In some cases, binary systems exhibit eclipses or occultations, and the timing of such events can be used to constrain the size of the two components. Knowledge of the mass and size of an object allows its density to be calculated, which in turn provides an indication of its composition and internal structure.

The discovery that 1998 WW31 is a binary system happened serendipitously — the authors were using a ground-based telescope to observe the object as part of a follow-up survey to assess the colour and position of a number of Kuiper-belt objects. The revelation that it is actually a binary system made it the first of seven such objects that have now been identified in the Kuiper belt. In fact, five of these binary systems were found during programmes designed to obtain colour or positional data on a large sample of objects. All of these discoveries were first reported in circulars issued by the International Astronomical Union. These circulars are used to describe initial observations, and are distributed to the astronomical community to encourage follow-up studies.

Both ground-based telescopes and the

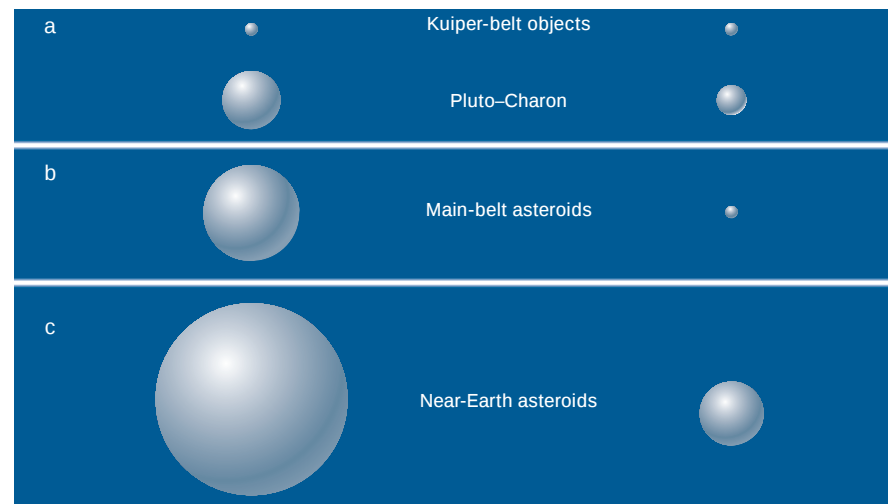


Figure 1 Comparison of typical binary systems within the Solar System. a, Kuiper-belt objects and the Pluto–Charon system (part of the Kuiper-belt system). b, Main-belt asteroids. c, Near-Earth asteroids. The separations are normalized to unity, but the sizes and separations of the two components are appropriately scaled in each panel. Some, but not all, of the diversity shown here may result from observational selection effects. Representative diameters for the larger bodies in each case are 100 km (Kuiper-belt objects), 2,300 km (Pluto), 100 km (main-belt asteroid) and 1 km (near-Earth asteroid); separation between the bodies is about 20,000 km, 8,000 km, 1,000 km and 2.5 km, respectively.

Hubble Space Telescope have played more-or-less equal roles in discovering Kuiper-belt binary asteroids, although ground-based observations dominate the discovery of main-belt binaries. Because of the angular-resolution limits caused by the Earth's atmosphere, the closest spacing between components of Kuiper-belt binaries that can be observed using ground-based telescopes is about one arcsecond (equivalent to the width of a coin seen from a distance of four kilometres). At the distance between the Earth and the Kuiper belt, which is roughly 40 times that between the Earth and the Sun, these angles mean that the minimum separation between the components that can be detected from the ground is about 20,000 km.

Both of the components in 1998 WW31 described by Veillet *et al.*¹ have diameters of about 100 km. Their orbital separation is more than a hundred times the size of the components (Fig. 1), so there can be barely any gravitational attraction between the two. Three of the binaries so far discovered using the Hubble Space Telescope feature larger components than those found in 1998 WW31, but despite their size, the distance between their elements is much shorter — about 8,000 km — and they are more reminiscent of the Pluto–Charon system (which is now generally taken more as an exceptional member of the Kuiper belt rather than a planet–moon combination). For the moment, it is hard to know whether such a division in the relationship between size and distance holds true throughout the population of Kuiper-belt binaries as there are so few examples.

Compared with samples of binaries in the main belt between Mars and Jupiter⁴, or in the population of binary objects with Earth-crossing orbits⁵, the orbital separations of Kuiper-belt binary objects observed to date are large (Fig. 1). Such wide separations between the components are truly arresting and defy accepted ideas about the processes of binary formation^{2–4}. For example, in the Earth–Moon system, tidal processes can explain the Moon's recession from the Earth at a present-day rate of about 3.8 metres per century, concomitant with a lengthening of the day of about 1.6 milliseconds per century. This results from conservation of total angular momentum in the system, with Earth's spin angular momentum being lost to orbital angular momentum. A simple calculation shows that 1998 WW31 and other widely separated Kuiper-belt binaries have too much angular momentum to have evolved in this manner. So they must have acquired their large orbital separations by some other mechanism.

Binary asteroids in the near-Earth population may form as a result of tidal disruption during close encounters with planets⁵, a mechanism that cannot explain the formation of main-belt or Kuiper-belt binaries.

Likewise, possible mechanisms of how main-belt binaries form are thought to produce only small separations between components, and cannot account for the large separations in Kuiper-belt binaries. So we have the disconcerting situation in which there is an abundance of binary systems in three distinct populations, with an apparent need for three separate mechanisms to account for how the systems are created.

The description of 1998 WW31 generates plenty of questions that need to be addressed. How did the system form? Why does it have such large orbital separation and angular momentum? How did it survive collisions? What does the large proportion of binary Kuiper-belt objects — estimated as at least 1% of the known population — indicate about the collisional environment in that population? The solutions to these

problems will improve our understanding of the collision and accretion processes that are key in making planets. They can be addressed by measuring the fundamental physical properties of large samples of objects, such as the binary systems and minor planets, that exist throughout our Solar System. ■

Jean-Luc Margot is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.
e-mail: margot@gps.caltech.edu

1. Veillet, C. *et al.* *Nature* **416**, 711–713 (2002).
2. Van Flandern, T. C., Tedesco, E. F. & Binzel, R. P. in *Asteroids I* (ed. Gehrels, T.) 466–479 (Univ. Arizona Press, Tucson, 1979).
3. Weidenschilling, S. J., Paolicchi, P. & Zappala, V. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 643–658 (Univ. Arizona Press, Tucson, 1989).
4. Merline, W. J. *et al.* in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) (in the press).
5. Margot, J. L. *et al.* *Science* online 11 April 2002 (DOI 10.1126/science.1072094).

Microbiology

A resistance switch

George A. O'Toole

Cystic-fibrosis patients often have drug-resistant *Pseudomonas aeruginosa* bacterial communities in their lungs. It seems that the bacterium's antibiotic resistance and ability to form communities are turned on together.

The vast majority of the world's bacteria exist not as free-living cells but as biofilms — structured communities that are attached to a surface. Biofilms can be formed by many different microorganisms on nearly any surface, from glass, metal and rocks to animal tissues and even Teflon. They are common in people who have become infected during a stay in hospital, and they colonize virtually every municipal water-pipe, possibly posing a threat to public health. Yet, despite such diversity, these microbial communities share several characteristics, including a complex three-dimensional architecture and increased resistance to antimicrobial drugs — a feature that makes them a persistent problem in clinical settings. On page 740 of this issue, Drenkard and Ausubel¹ provide clues to how one such bacterium, the opportunistic pathogen *Pseudomonas aeruginosa*, regulates its shift to the antibiotic-resistant state.

Antibiotic-resistant biofilms develop in several different contexts. They are particularly evident on infected implants. Any foreign surface in a human body — from catheters and artificial hips to contact lenses — is a potential site of bacterial colonization, and biofilms on such surfaces tend to be recalcitrant to standard antimicrobial treatments, often requiring removal of the implant. The economic cost of these infections is an estimated \$6 billion a year in the United States alone². And antibiotic-resistant biofilms may be involved in other types

of infection, too. For example, it is likely that *P. aeruginosa* grows as biofilms in the lungs of patients with cystic fibrosis, and it is thought that the near-impossibility of eliminating this bacterium results, at least partly, from biofilm-specific resistance to standard antibiotic treatments³. In addition, biofilms of *Haemophilus influenzae* might be involved in otitis media (better known to parents simply as earache), perhaps explaining why *H. influenzae* infections recur after antibiotic treatment has ceased⁴.

Most discussions of microbial antibiotic resistance include the idea that bacteria mutate in a way that renders them resistant to one or more antibiotics. Alternatively, they can acquire from other microorganisms whole DNA molecules or mobile genetic elements that, for example, encode proteins that pump out or destroy antibiotics. But these mechanisms usually relate to free-living bacteria in suspension (planktonic bacteria). In contrast, biofilm-specific antibiotic resistance is thought to arise because the cells enter a new physiological state: biofilm bacteria are genetically identical to their planktonic counterparts — they have not acquired new mutations or genetic elements — and once they return to planktonic growth, the biofilm-specific resistance is lost⁵. It is unclear how biofilms become resistant, although various explanations have been proposed, including limitations to the diffusion of antibiotics because of biofilm architecture; the activation of an

inherent response to stress or starvation; or the increased expression of pump proteins (reviewed in ref. 6).

Drenkard and Ausubel¹ tackled the problem of the antibiotic resistance of biofilms by studying *P. aeruginosa*. They first cultured a *P. aeruginosa* strain on agar plates containing one of several medically useful antibiotics, such as tobramycin, gentamycin or kanamycin. Antibiotic-resistant *P. aeruginosa* colonies developed, many of which were small and had rough edges (and so were dubbed 'rough small-colony variants'). Compared with the original strain, these variants also formed larger biofilms, and more rapidly. These observations imply that biofilm formation and antibiotic resistance are switched on together.

A similar situation may occur in cystic-fibrosis patients, as the authors showed by studying the colonies that arise after culturing sputum samples from patients' lungs. All the *P. aeruginosa* samples from patients who were being treated with antibiotics were small-colony variants, and many were highly resistant even to antibiotics not being used for treatment. (Resistance to a wide range of antimicrobial drugs is another characteristic of biofilms.) So the antibiotic-resistant, small-colony variants seem to be important in cystic-fibrosis patients, as well as *in vitro*. The authors found few or no small-colony variants in untreated patients.

How might the switch to antibiotic resistance and biofilm architecture be controlled? The small-colony variants derived from the original *P. aeruginosa* strain reverted to the original characteristics (phenotype) when cultured without antibiotics¹. This suggests that temporary phenotypic changes are involved, rather than permanent genetic alterations. More specifically, these changes could be due to 'phase variation' — a mainly random process whereby bacteria use several molecular mechanisms to express phenotypes that may be advantageous in a particular environment. The process is not typically controlled by the environment, and involves randomly producing many different phenotypic variants, the goal being one that can survive and proliferate.

But there are cases of environmental signals specifically controlling the rate of phenotypic switching⁷, and one well-studied example involves the expression by *Escherichia coli* of surface structures called Pap pili, which enable the bacterium to attach to its host. At 37 °C (body temperature) the apparently random addition of methyl groups to regulatory sequences flanking the *pap* genes dictates whether the expression of Pap pili is on or off. But changes in temperature are one environmental stimulus that affect this on/off switch, and the expression of Pap pili is completely shut down at 26 °C or below⁸ (presumably when *E. coli* has left its host

and no longer needs these structures).

Similarly, Drenkard and Ausubel's results¹ suggest that environmental cues — such as the conditions in the lungs of cystic-fibrosis patients — may control the degree of antibiotic resistance of *P. aeruginosa*. They identified a *P. aeruginosa* regulatory protein that, when overexpressed or mutated, altered the rate at which rough small-colony variants reverted to the antibiotic-sensitive phenotype. The authors propose that this protein is not directly involved in the modifications required for phenotype switching, but may regulate the factors that are required.

So Drenkard and Ausubel have described a general mechanism — environment-driven phenotypic switching — that explains how antibiotic-resistant biofilm variants of *P. aeruginosa* arise. They also propose that these particular variants account for the antibiotic resistance of *P. aeruginosa* in cystic-fibrosis patients. Previous work supports this contention: it seems that not all biofilm bacteria are highly antibiotic resistant. Indeed, most biofilm cells die after treatment with antibiotics, but a small population survives, and is eliminated only when high antibiotic concentrations are used^{9,10}. Moreover, it has been proposed that some biofilm cells develop a phenotype that renders them resistant to antimicrobial drugs^{11,12}, and that this developmental programme is triggered by environmental signals. The highly resistant cells are called persisters^{12,13}. Perhaps the small-colony variants identified by Drenkard and Ausubel¹ are likewise persisters.

We still do not know what characteristics enable these cells to survive in the face of antibiotics. But it might be possible to develop treatments that target persisters — for example, by modifying the regulatory protein of *P. aeruginosa* — thereby rendering remaining biofilm cells susceptible to conventional antibiotics. ■

George A. O'Toole is in the Department of Microbiology and Immunology, Room 202, Vail Building, Dartmouth Medical School, Hanover, New Hampshire 03755, USA.
e-mail: georgo@Dartmouth.edu

1. Drenkard, E. & Ausubel, F. M. *Nature* **416**, 740–743 (2002).
2. Archibald, L. K. & Gaynes, R. P. *Nosocomial Infect.* **11**, 245–255 (1997).
3. Singh, P. K. et al. *Nature* **407**, 762–764 (2000).
4. Rayner, M. G. et al. *J. Am. Med. Assoc.* **279**, 296–299 (1998).
5. Hoyle, B. D. & Costerton, W. J. *Prog. Drug Res.* **37**, 91–105 (1991).
6. Mah, T.-F. & O'Toole, G. A. *Trends Microbiol.* **9**, 34–39 (2001).
7. Hallet, B. *Curr. Opin. Microbiol.* **4**, 570–581 (2001).
8. White-Ziegler, C. A., Angus Hill, M. L., Braaten, B. A., van der Woude, M. W. & Low, D. A. *Mol. Microbiol.* **28**, 1121–1137 (1998).
9. Muli, F. W. & Struthers, J. K. *J. Med. Microbiol.* **47**, 401–405 (1998).
10. Ashby, M. J., Neale, J. E., Knott, S. J. & Critchley, I. A. *J. Antimicrob. Chemother.* **33**, 443–452 (1994).
11. Cochran, W. L., Suh, S. J., McFeters, G. A. & Stewart, P. S. *J. Appl. Microbiol.* **88**, 546–553 (2000).
12. Lewis, K. *Antimicrob. Agents Chemother.* **45**, 999–1007 (2001).
13. Stewart, P. S. & Costerton, J. W. *Lancet* **358**, 135–138 (2001).

Daedalus

Very slow vacuum

Fast-moving particles or atoms leave tracks in many materials. Geologists assess the ages of specimens by studying the tracks of particles or recoil nuclei; microscopists use molecules or ions like light, firing them at surfaces or through samples to gain information. Technologists can make many films: polyalkane, polyester, polyamide, their laminates, and so on. DREADCO physical chemists are now firing atoms, molecules and ions through thin films from one side only. Most of the resulting damage-tracks will just be fine holes. But Daedalus reckons that, every so often, a molecular group will be pushed away but not detached. Even if it returns to its place it will easily give way again, in that same direction. He is looking for a molecular one-way valve, the equivalent of the rectifier in electronics.

Initial progress must depend on luck. Daedalus cannot guess what molecule or ion, fired at what energy at what film, will give the best one-way valve for what gas molecules. But the resulting DREADCO 'Vacfilm' will have many uses, especially for the molecules in air. Seal something inside it; if its pumping direction is outwards, it will slowly extract all internal air. The Maxwell–Boltzmann distribution gives some molecules twice the average energy or even more; these are the ones that will pass the little valves. But the process may take many weeks. Daedalus reckons that Vacfilm, like an inefficient Carnot engine, will be driven by small daily changes of temperature. Conversely, however, a mere atmosphere of adverse pressure is nothing to a molecular valve. In due course vacuum will be achieved and maintained; even small leaks will be pumped out.

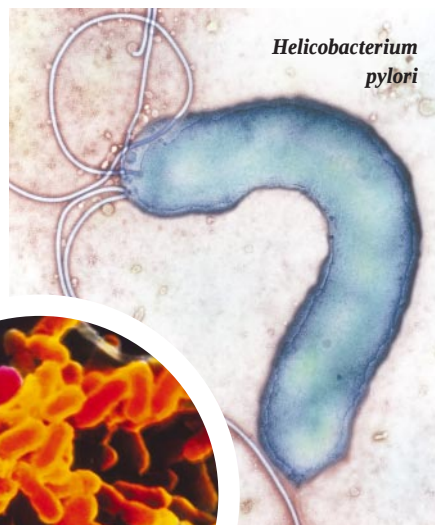
The obvious use is food wrap. Vacuum packing is already used for products such as bacon. Many products would last almost indefinitely if protected from oxygen, now imperfectly excluded by aluminized film. A self-pumping food film would be widely welcomed. But Daedalus also recalls the refrigerator. It uses almost all its power pumping away heat, which just flows back through bad insulation. The best foamed insulation should be filled with vacuum, yet air always slowly leaks in. A block of foam inside sealed Vacfilm should be ideal. It would keep the heat out indefinitely, and any small air-leaks would soon be pumped away. The constant temperature-swings of the refrigerator thermostat would keep the Vacfilm working.

David Jones

Microbial genomes multiply

Russell F. Doolittle

It is seven years since the first bacterial genome was completely sequenced, and more than 60 others have now been determined. What has been the impact of these projects on pure science and public welfare?



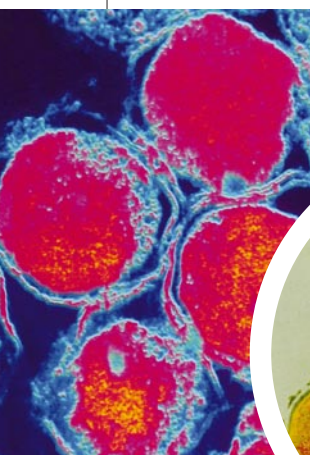
*Helicobacterium
pylori*



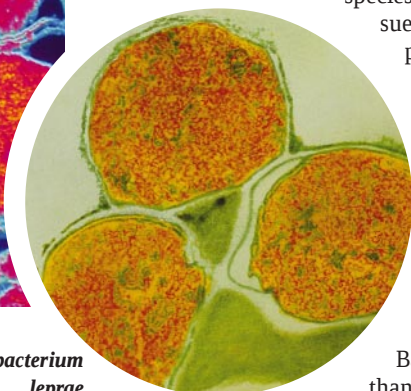
*Rickettsia
prowazekii*



*Mycobacterium
tuberculosis*



*Chlamydia
trachomatis*



*Mycobacterium
leprae*

The publication of the first complete sequence of a bacterial genome¹ in 1995 was a signal event, underscored by the fact that the article has been cited more than 2,100 times during the intervening seven years. It was a marvellous technical achievement, made possible by automatic DNA-sequencing machines. The feat is the more impressive in that complete genome sequencing has now been adopted in many different laboratories around the world. Four years ago in these columns I examined the situation after a dozen microbial genomes had been completed². Now, with upwards of 60 microbial genome sequences determined and twice that many in progress, it seems reasonable to assess just what is being learned. Are new concepts emerging about how cells work? Have there been practical benefits in the fields of medicine and agriculture? Is it feasible to determine the genomic sequence of every bacterial species on Earth? The answers to these questions may be Yes, Perhaps and No, respectively.

The whole-genome sequence era must be viewed against a backdrop of 100 years or more of biochemistry and bacteriology. Although there have been some revelations of late, much of what is being found was fully expected. Still, the wholeness itself is a tremendously valuable asset, as well as all the details: everything is on the table, the entire parts list — well, as we will see, almost everything.

Beyond the matter of completeness, comparisons of genomes are the main attraction. In the case of medically important pathogens, comparisons between strains can pinpoint differences between the virulent and the avirulent, and comparisons between species can be informative about host or tissue specificity. At the generic level, comparisons begin to reveal the fundamental divergences of different microbial ways of life and their evolutionary origins.

Microbes encompass all three realms of living organisms, including unicellular eukaryotes (organisms with well-defined nuclei and cytoskeletons). When it comes to size, the prokaryotic genomes from the Archaea and the Bacteria sequenced so far span more than an order of magnitude, from the

mere 600,000 base pairs (bp) of some mycoplasmas (now called mollicutes) to almost 8 million for the nitrogen-fixing root-nodule bacterium, *Mesorhizobium loti* (Fig. 1, overleaf). By comparison, the genomes of eukaryotes — several of which are completed, with many others being explored — range from less than 3 million bp for an intracellular microsporidian to the 4 billion found in the human genome, and more.

The genomes of eukaryotes are, however, often greatly inflated by the presence of considerable amounts of non-coding DNA, including both intergenic and intragenic (intron) sequences. In contrast, most prokaryotes have their genes tightly packed together with very little intergenic space, and they do not have introns in the genes that encode proteins. When the genes themselves are counted, the information content of prokaryotes and eukaryotes is not nearly so disparate as would seem from raw genome size. Indeed, the number of genes in the largest bacterial genomes actually exceeds the number in some eukaryotes. For example, *M. loti* has about 8,000 genes³, but *Saccharomyces cerevisiae* (baker's yeast) has only 6,200 (ref. 4), and the microsporidian *Encephalitozoon cuniculi* only about 2,000 (ref. 5)!

Sequencing strategy

The method of whole-genome shotgun sequencing has been used for all the microbial genomes. Shotgun sequencing is not a new concept: in fact, it was the conventional method for sequencing proteins almost half a century ago. The strategy is to fragment an informational polymer such as a protein or DNA molecule into large pieces, determine partial sequences, and then put the fragments in order by finding overlapping regions that have identical sequences at opposite termini. The breakthrough in whole-genome sequencing occurred when it became possible to 'read' a length of around 500 consecutive bases on a single, cloned fragment of DNA, at which point the arithmetic of how many random fragments would be needed to reconstruct the entire order of a million or more bases showed that a practicality threshold had been reached. The entire 1.83 megabase sequence of the bacterium *Haemophilus influenzae* was determined by sequencing fewer than 20,000

fragments, with the added advantage that most positions were confirmed by multiple determinations¹.

Once the overall DNA sequence has been reconstructed by the appropriate computer programs, the important job of finding the genetic information begins. A small portion of the genome codes for various RNA structures, including the ribosomal and transfer RNAs, and these are easily spotted. Most of a bacterial genome is devoted to genes for proteins. These regions are identified by computer software that first translates the DNA sequence into amino acids according to the rules of the genetic code, in which three-base DNA 'triplets' (or the corresponding codons in messenger RNA) each encode an amino acid. Some DNA triplets correspond to the punctuation marks that delineate genes, including the initiation codon at the start and the termination codon ('stop sign') at the end of a gene. If there is a significant uninterrupted run of amino-acid codons — say 50 or more — between an initiation and a termination codon, the DNA segment is presumed to encode a protein in that reading frame, and the region is referred to as an open reading frame, or ORF.

After the ORFs are tallied up, it remains to find out what they encode. This phase of the operation involves computer searches of large databases of known protein sequences. Significant matches are evaluated, and judgements are made about whether an ORF might encode exactly the same function as the protein retrieved from the database, or a related function. The judgement depends on two factors: the degree of similarity between the sequences (the new ORF and those in the database), and the closeness of the relationship of the organisms from which they were obtained. It is not a foolproof method of ascertaining gene function, and is often complicated by horizontal gene transfers, about which more below.

Choosing the targets

Although medically important bacteria account for only a tiny fraction of the Earth's bewildering array of prokaryotes, three-quarters of the completely sequenced members of the Bacteria are of clinical significance. Despite a half-century of 'miracle-drug' antibiotics, bacterial diseases remain a scourge. As such, humanitarian and commercial forces are allied in the hope of finding genetically controlled chinks in the bacterial armour that can be exploited by new custom-designed drugs, or discovering gene products that can aid in the development of vaccines and diagnostic tools.

Even though most reports of these bacterial genome sequences begin with a litany of how menacing the bacteria are, many were chosen for other reasons as well. For exam-

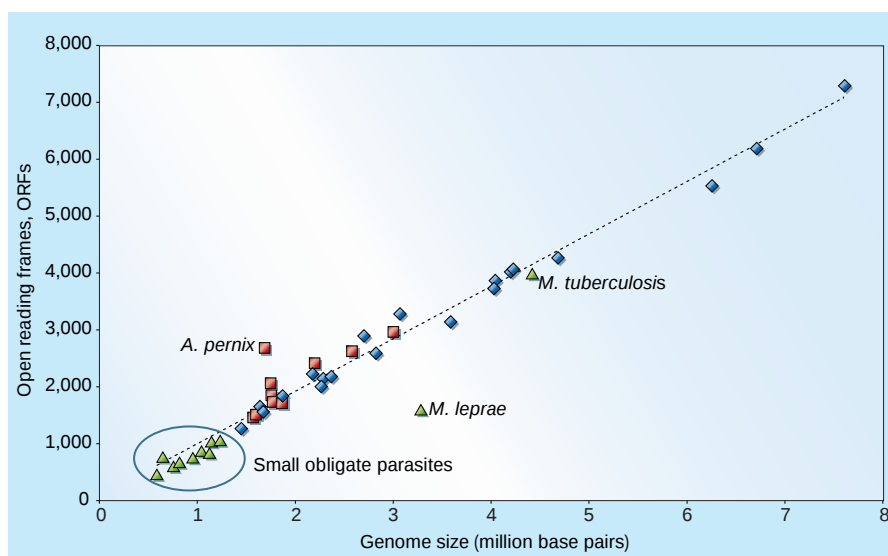


Figure 1 Number of genes (ORFs) plotted against genome size for 44 fully sequenced genomes, including ten Archaea (squares) and 34 Bacteria. Obligate bacterial parasites are denoted by triangles; all other bacteria are shown as diamonds. *Mycobacterium leprae* is a genome 'in decay' that has a large number of pseudogenes. The archaeon *Aeropyrum pernix* is unusual in having an excessive number of duplicated ORFs.

ple, the small intracellular bacterium *Rickettsia prowazekii* is the cause of typhus, but it also has a great bearing on understanding how mitochondria evolved⁶. In addition, many of the genomes initially selected for study were conveniently small, and thus appropriate for the developing technology.

Not all the newly completed genomes come from medically important bacteria. Some were targeted because of their agricultural importance, and others have potential for bioremediation — the clean-up of polluted sites — or other ecological applications. Several of the Archaea are methanogens that produce most of the world's atmospheric methane, a 'greenhouse gas'. Although there are no known pathogens among the Archaea, there is great commercial interest in the heat-loving thermophiles among them, based on the hope that thermostable enzymes useful to industry will emerge.

Remarkably, the nonpatentable intellectual fallout from these studies has been enormous, perhaps at present outweighing the medical or practical gain. This is partly because at this stage of our knowledge the genome of any organism is a treasure trove for the biologist. Imagine knowing every genetically determined component of an organism and what it does! That coveted goal is not yet ours, however, mostly because in each of the completely sequenced genomes so far there are vast numbers of putative genes for proteins of unknown function.

Genome surprises

It is precisely this point — the large numbers of putative genes with no known function — that has been the biggest surprise in genome sequencing. The genome of the archaeon

Aeropyrum pernix, for example, contains more than 1,500 ORFs — 57% of its total gene content — not recognizable by computer searching in any other organism⁷. And more than 40% of the approximately 4,000 ORFs found in *Mycobacterium tuberculosis*, one of the best-studied bacteria of the past century, fall into the same category⁸. In every genome examined so far, at least a quarter of the genes remain 'hypothetical', in that no function can be ascribed. After such a long history of biochemical and genetic examination, how could there be so much in the way of unknown equipment?

The hypothetical ORFs fall into two categories: those that are found in a variety of organisms, and which almost certainly encode functional proteins; and those that are unique to particular lineages. The latter can sometimes be attributed to runaway gene duplication; many of the unidentified genes in *A. pernix* are in this category⁷. The extra ORFs in this case are unusually small, pushing the ratio of number of genes-to-genome size high above expectation (Fig. 1); not all of them may encode proteins. In contrast, there are large numbers of unidentified genes in a variety of organisms that look conventional in every way. Where these unique sequences are coming from and what they do remain baffling mysteries.

Lateral gene transfers

Although the transfer of genetic material between distantly related organisms was well known before the whole-genome sequence era, the magnitude of the exchange that has occurred between different prokaryotes — including numerous gene exchanges between the realms of Bacteria and Archaea — was certainly unappreciated. Horizontal

transfers are ordinarily discovered during the construction of a phylogenetic tree of an individual protein. When two sequences from otherwise distantly related organisms are found to be more similar to each other than pairs of sequences from known closer relatives, horizontal gene transfer is suspected. In the new world of genomics, however, the list of potential horizontal transfers is more often compiled during database comparisons of newly found ORFs. If a match score for a protein in a distantly related organism is higher than that for the protein from a nearer relative, transfer is automatically presumed. The method has its weaknesses, however, as detailed in ref. 9.

Mechanisms for horizontal exchange in the prokaryotic world are well known. Bacterial viruses (bacteriophages) can move genes from one species of bacterium to another by the process of transduction, whereas the more direct movement of naked DNA by transformation commonly involves plasmids.

Plasmids are autonomously replicating elements composed solely of DNA. They are usually — but not always — circular; large ones are sometimes considered mini-chromosomes. Unlike the main bacterial chromosome, there may (or may not) be many copies of a plasmid within an individual cell. Plasmids often contain genes that, from the cell's point of view, may be 'unessential but desirable'. The best-known examples are genes whose products are used to disable antibiotics, but plasmid-borne traits include a host of other attributes.

The plasmid repertoire is maintained by the lottery that accompanies plasmid replication. If a bacterium divides and plasmid replication does not keep pace, one daughter cell may end up without a plasmid. If conditions are not threatening, the bereft daughter and her progeny may thrive anyway, but if the environment is hazardous, the plasmid-bearers will be the ones to carry on.

In prokaryotes, genes are often bunched on the chromosome by function, and — not

unexpectedly — these clusters can move from species to species as a group¹⁰. Microbial genomes are constantly being scrambled by the cutting and pasting that accompanies transposition (the movement of DNA sequences to other positions on the chromosome) and other recombinational activities. The driving force responsible for gene clustering may in fact be the potential for gene spread to other organisms, the logic being that moving a single gene without the genes that interact with it has a high probability of being a dead end¹¹.

As such, it is hardly surprising that clusters of genes are transferred from species to species on plasmids. Nonetheless, the very large size and widespread frequency of some of these gene clusters is astonishing, and the bearing on pathogenicity is enormous.

What makes a pathogen?

There are pathogenic and nonpathogenic strains of most medically important bacteria, and virulence factors were, of course, known and characterized long before the advent of the genomic projects. In 1944, the classic experiment of Avery, MacLeod and McCarty¹² involved the transformation of a non-encapsulated, avirulent strain of pneumococcus (now called *Streptococcus pneumoniae*) with DNA from a capsulated, virulent strain. It was presumed that the capsule itself played a significant part in virulence, protecting the bacterium from the host's defence systems. In confirmation, the complete genome sequence revealed that a cluster of a dozen genes in *S. pneumoniae* is needed for capsule synthesis¹³. Although the complete genome sequence of the avirulent strain has not yet been reported, microarray hybridization studies have shown the cluster to be absent, confirming that the capsule is the primary virulence factor in these bacteria¹³.

Many other genetically controlled characters can confer virulence on bacteria besides encapsulation, including factors that enable the bacteria to attach to and disrupt host cells.

Many pathogenic bacteria have a similar machinery for injecting proteins into the cytoplasm of the host's cells, and the sophisticated wherewithal to do so seems to be exchanged among them frequently. One of these amazing 'machines', the type III secretion system¹⁴, has been identified in a variety of fully sequenced bacteria, ranging from the tiny genome of the bacterium *Chlamydia trachomatis*¹⁵, an intracellular animal parasite, to the enormous genome of *M. loti*³.

Pathogenicity islands¹⁶ are very large gene clusters on bacterial chromosomes that are highly correlated with virulence. As an example, pathogenic strains of the bacterium *Helicobacter pylori*, which has been implicated in the formation of gastric ulcers, have a 40,000-bp 'island' of DNA that includes a large number of genes involved in attacking host cells¹⁶. Avirulent strains of this bacterium lack this region. Although the concept of pathogenicity islands was developed well before whole-genome projects began, it has been greatly illuminated by the new sequence data. In *Escherichia coli* and other gamma-proteobacteria such as *Salmonella*, clusters of genes with functions for aiding and abetting the disruption of host cells are delineated in the chromosome by 'insertion sequences' or other sequences characteristic of transposable elements. Among these are transfer RNA genes, which may serve as targets for special excision and integration enzymes. On occasion, the entire island may be cut out and moved to a plasmid. Once on a plasmid, gene clusters can migrate to other bacteria where they can be reintegrated in a new genome¹⁶.

Pathogenicity islands have also been implicated in strain differences in *Pseudomonas aeruginosa*¹⁷, which often infects humans. Strains of *P. aeruginosa* can be divided into two types, a and b, and only type a has flagellar proteins that are glycosylated. Genomic comparisons of the two types revealed that type a has a 16-kilobase (kb) island containing 14 genes of the sort that synthesize and assemble sugars, embedded right in the middle of an even larger gene cluster known to be responsible for the manufacture of flagella. Remarkably, the b type has a different, smaller, island of three genes of unknown function at exactly the same location. Apparently the two cassettes — one composed of 14 genes, the other of three — can be exchanged between strains by reciprocal recombination¹⁷.

In yet another case, recently in the lime-light for unhappy reasons, *Bacillus anthracis* (the cause of anthrax) contains two large plasmids, one of which has a 44.5-kb pathogenicity island¹⁸. This island, which contains genes for — among other things — the toxin that can be so lethal to humans, is flanked by inverted insertion sequences. The plasmid also contains a collection of what seem to be transposases and integrases, suggesting a

Organism (number of genes)	Glycolysis	Tricarboxylic- acid cycle	Amino-acid biosynthesis	Purine biosynthesis	Pyrimidine biosynthesis	Ancestral stock
<i>Mycoplasma genitalium</i> (470)	+	–	–	–	–	<i>Bacillus–Clostridium</i>
<i>Buchnera</i> species (588)	+	–	+	+	+	Gamma- proteobacteria
<i>Rickettsia prowazekii</i> (834)	–	+	–	–	–	Alpha- proteobacteria
<i>Chlamydia trachomatis</i> (894)	+	–	+	–	–	Main line
<i>Treponema pallidum</i> (1,041)	+	–	–	–	–	Main line
<i>Mycobacterium leprae</i> (1,604)	Partial	In decay	+	+	+	<i>Bacillus–Clostridium</i>

Figure 2 **Many routes to intracellular adaptation.** The differing presence (+) or absence (–) of certain metabolic pathways in the streamlined genomes of parasitic bacteria shows how variable the process may be.

history of shuffling and exchange. Interestingly, the sequence of the main chromosome of *B. anthracis* — still on the unfinished list — is exceedingly similar to those of *B. cereus* and *B. thuringiensis*, neither of which carry the plasmids¹⁹. *B. thuringiensis*, which produces a toxin fatal to lepidopteran caterpillars, is thought to be the most commonly used biological pesticide worldwide.

Not all genomic islands encode genes for pathogenicity. In *M. loti*, for example, the main chromosome has a 'symbiotic island' of more than 600,000 bp, which is necessary for the bacterium to establish a symbiosis with its legume host plant. It is flanked by 17-bp repeats and has a codon usage that is significantly different from the rest of the chromosome, which suggests an 'alien' origin³. The symbiotic island contains 580 genes — more than some small bacterial genomes in their entirety. Among them are several dozen genes that encode proteins involved in nodulation and nitrogen fixation. That they have been introduced into the chromosome from an ancestral plasmid seems certain, in that another root-nodule bacterium, *Sinorhizobium meliloti*²⁰, has a much smaller main chromosome and two very large plasmids, one of which contains most of the genes that make up the island on the main chromosome of *M. loti*, including the corresponding genes for nodulation and nitrogen fixation.

Adaptive gene losses

One of the most fascinating phenomena to emerge from the microbial genome studies so far is the extent to which parasitic bacteria have adapted to life in animal hosts. It is not only that many of these organisms have lost numerous genes as they became dependent on materials supplied by their hosts. Rather, it is that the very process of losing those genes has been captured in the moment. In the case of the typhus bacterium *R. prowazekii*, almost 25% of the genome is non-coding⁶, in contrast to the 10% non-coding DNA that typifies most bacteria. Some of the non-coding DNA actually corresponds to pseudogenes, segments that are still recognizable as having encoded proteins in the past, but which now contain stop codons and/or deletions that keep the gene from being expressed properly.

An even more dramatic illustration of rampant gene loss is afforded by the genomic sequence of *Mycobacterium leprae*²¹. The relatively short period during which gene decay has been going on in this bacterium is underscored by the fact that fully half of its still relatively large genome is non-coding. More than 1,100 pseudogenes were uncovered. Eventually these non-functional genes will disappear from the genome, the result of random deletion. Having the complete genome of the closely related *M. tuberculosis* for comparison, a genome that is clearly not

in the process of immediate genetic decay⁸, affords a unique view of where the *M. leprae* genome is coming from.

Most of the 1,100 pseudogenes in *M. leprae* are apparently fully functional ORFs in *M. tuberculosis*. Remarkably, those genes of *M. leprae* that are still intact have a very high sequence resemblance to those of *M. tuberculosis*, implying that these two species have had a common ancestor quite recently, perhaps only a few million years back. What triggered this massive gene decay in *M. leprae*? Will *M. tuberculosis* follow suit?

Do bacteria with streamlined genomes always lose the same genes? The answer is equivocal. When it was found initially that both *Borrelia burgdorferi*, the agent responsible for Lyme disease, and *Mycoplasma genitalium*, a parasite of the urogenital tract, had both lost the capacity to synthesize essential amino acids, it was viewed as a kind of 'convergent evolution'²². But as more of these reduced genomes are sequenced, it has become apparent that the process is idiosyncratic: there are many routes to intracellular adaptation (Fig. 2).

Some of these parasites, like *Mycoplasma* and *Borrelia*, depend on the host to provide key metabolites. Others, like *Rickettsia* and *Chlamydia*, are energy parasites that literally steal ATP from the host cell. Even though these latter two bacteria are very distantly related, they do this with the same alien enzyme, an ADP/ATP translocase of a sort not found in the host-cell mitochondria or, so far as is known, any place elsewhere in the bacterial world. Until very recently, the only known relatives of the bacterial version were found in chloroplasts²³. How did these two widely divergent bacteria, both animal parasites, happen upon the same strategy with the same enzyme, and how and where did they acquire it? Compounding the mystery, the same protein was found to be encoded in the recently sequenced genome of the tiny eukaryote *E. cuniculi*, itself an obligate intracellular pathogen⁵. Presumably the parasite uses the protein to scavenge ATP from its host cell, just like the small bacterial pathogens.

The streamlining that comes with parasitism is not only associated with disease. In fact, the vast majority of instances in which bacteria live inside eukaryotic cells are mutually profitable to both host and guest. Consider the fascinating case of *Buchnera*, a relative of *E. coli* that is now an obligate parasite of aphids, and whose relatively small 640-kb genome sequence was recently completed²⁴. Neither the aphid nor the bacterium can live without each other. In this case, the bacterial symbiont provides essential nutrients to the host cell, often using host-cell metabolites as starting materials. In exchange, the bacterium has gained a safe haven and has come to depend on the insect cell to provide vital materials, including even its cell membrane. Left to its own devices, would *M. leprae* evolve

so that a similar mutually beneficial situation could occur in human cells? Obviously we can't afford to wait and see.

Outlook

Meanwhile, hopes for biotechnological innovation remain high, even though the technical problems associated with translating all the new information into action have been unexpectedly resistant to solution²⁵. Some of the most interesting genes have proved difficult to express and purify when engineered into conventional *E. coli* hosts. And the matter of all those ORFs whose functions remain unknown is a serious roadblock to a genuine understanding of cellular events.

It seems to me that, although evolutionists have been the unintended beneficiaries of many of these medically directed genome projects, in the long run the medically motivated — whatever their reasons — could profit greatly from the insights of the evolutionists, if the latter are given more voice in the selection of genomes to sequence. There are myriads of microbes on Earth²⁶, and it is out of the question to think about sequencing them all. A truly representative set is needed if we are to gain a proper perspective. In this regard, one of the most sinister items to appear in this journal in recent times reported that the Bush administration is considering "investment criteria" for basic research²⁷. Such a policy would be likely to doom the mutually profitable interplay between scientific curiosity and the public good. ■

Russell F. Doolittle is in the Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093-0634, USA.
e-mail: rdoolittle@ucsd.edu

1. Fleischmann, R. D. *et al.* *Science* **269**, 496–512 (1995).
2. Doolittle, R. F. *Nature* **392**, 339–342 (1998).
3. Kaneko, T. *et al.* *DNA Res.* **6**, 331–338 (2000).
4. Goffeau, A. *et al.* *Science* **274**, 546–567 (1996).
5. Katinka, M. D. *et al.* *Nature* **414**, 450–453 (2001).
6. Andersson, S. G. E. *et al.* *Nature* **396**, 133–140 (1998).
7. Kawarabayashi, Y. *et al.* *DNA Res.* **6**, 145–152 (1999).
8. Cole, S. T. *et al.* *Nature* **393**, 537–544 (1998).
9. Koski, L. B. & Golding, G. B. *J. Mol. Evol.* **52**, 540–542 (2001).
10. Crawford, I. P. & Milkman, R. in *Evolution of the Molecular Level* (eds Selander, R. K., Clark, A. G. & Whittam, T. S.) 77–95 (Sinauer, Sunderland, Massachusetts, 1991).
11. Lawrence, J. G. & Roth, J. R. *Genetics* **143**, 1843–1860 (1996).
12. Avery, O. T., MacLeod, C. M. & McCarty, M. *J. Exp. Med.* **79**, 137–158 (1944).
13. Tettelin, H. *et al.* *Science* **293**, 498–506 (2001).
14. Lee, C. A. *Trends Microbiol.* **5**, 148–156 (1997).
15. Stephens, R. S. *et al.* *Science* **282**, 754–759 (1998).
16. Hacker, J., Blum-Oehrl, G., Muhlendorfer, I. & Tschape, H. *Mol. Microbiol.* **23**, 1089–1097 (1997).
17. Arora, S. K., Bangera, M., Lory, S. & Ramphal, R. *Proc. Natl. Acad. Sci. USA* **98**, 9342–9347 (2001).
18. Okinaka, R. T. *et al.* *J. Bacteriol.* **181**, 6509–6515 (1999).
19. Helgason, E. *et al.* *Appl. Environ. Microbiol.* **66**, 2627–2630 (2000).
20. Galibert, F. *et al.* *Science* **293**, 668–672 (2001).
21. Cole, S. T. *et al.* *Nature* **409**, 1007–1011 (2001).
22. Fraser, C. M. *et al.* *Nature* **390**, 580–586 (1997).
23. Mohlmann, T. *et al.* *Eur. J. Biochem.* **252**, 353–359 (1998).
24. Shigenobu, S. *et al.* *Nature* **407**, 81–86 (2000).
25. Weinstock, G. M., Smajs, D., Hardham, J. & Norris, S. J. *Res. Microbiol.* **151**, 151–158 (2000).
26. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. *Proc. Natl. Acad. Sci. USA* **95**, 6578–6583 (1998).
27. Macilwain, C. *Nature* **413**, 5 (2001).

A mutant gibberellin-synthesis gene in rice

New insight into the rice variant that helped to avert famine over thirty years ago.

The chronic food shortage that was feared after the rapid expansion of the world population in the 1960s was averted largely by the development of a high-yielding semi-dwarf variety of rice known as IR8, the so-called rice 'green revolution'¹⁻³. The short stature of IR8 is due to a mutation in the plant's *sd1* gene, and here we identify this gene as encoding an oxidase enzyme involved in the biosynthesis of gibberellin, a plant growth hormone. Gibberellin is also implicated in green-revolution varieties of wheat, but the reduced height of those crops is conferred by defects in the hormone's signalling pathway⁴.

There are various reasons for the dwarf phenotype in plants, but gibberellin (GA) is one of the most important determinants of plant height⁵⁻⁷. To investigate whether the *sd1* gene in semi-dwarf rice (Fig. 1a) could be associated with malfunction of gibberellin, we tested the response of this mutant to the hormone. We found that *sd1* seedlings are able to respond to exogenous gibberellin, which increases their height to that of wild-type plants (results not shown).

Gibberellin is synthesized from geranylgeranyl diphosphate in higher plants, with an aldehyde intermediate being converted by a sequence of oxidase-catalysed reactions to a series of gibberellin precursors (designated here by GA subscripts). We found that the GA₂₀ intermediate was depleted in the *sd1* mutant relative to the wild type, but that GA₅₃ (produced earlier in the pathway) was accumulating (results not shown). These results indicate that the activity of GA₂₀ oxidase (GA20ox), a key enzyme in the biosynthesis of gibberellin that catalyses the three steps GA₅₃→GA₄₄→GA₁₉→GA₂₀, is not functioning effectively in the mutant.

A gene encoding a GA20ox isoenzyme (GA20ox-1) has been isolated from rice⁸, but this does not correspond to the *sd1* locus (results not shown). However, we isolated a new GA20ox gene (GA20ox-2) by using degenerate primers based on the conserved domain of the GA20ox genes in rice (GA20ox-1)⁸ and *Arabidopsis* (GA5)⁹, and found that GA20ox-2 was located on the long arm of chromosome 1, tightly linked to the *sd1* locus¹⁰ (Fig. 1b).

The deduced amino-acid sequence of GA20ox-2 showed 47.8% identity to GA20ox-1 and 49.5% identity to *Arabidopsis* GA5 (results not shown). When we compared the GA20ox-2 gene sequences from four *sd1* mutants to that in the wild type, we found that one *sd1* allele contains a 383-base-pair deletion (the semi-dwarf rice strain 'dee-geo-woo-gen' and IR8 both

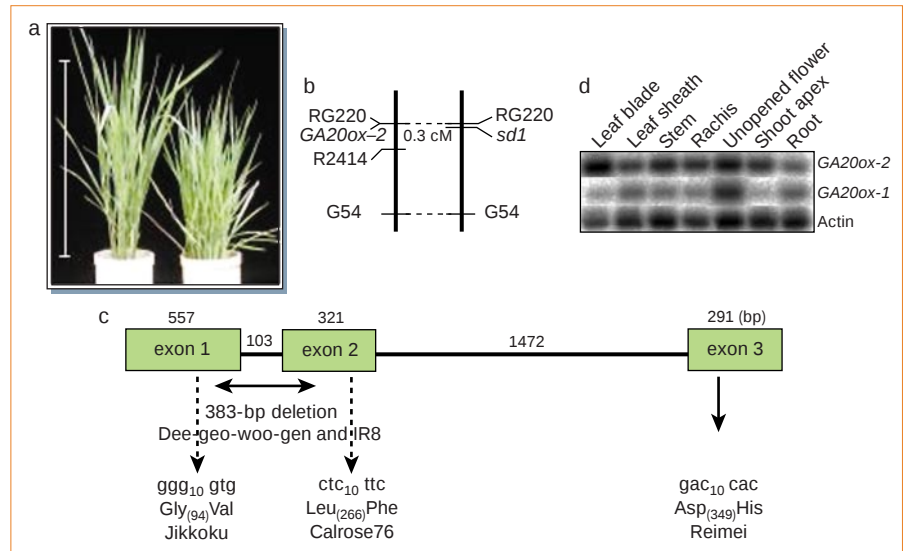


Figure 1 Effect of a mutant gibberellin-biosynthesis gene in rice. **a**, Morphology of *sd1*-mutant rice plants: left, taichung 65 (wild type); right, IR8 (*sd1*). Scale bar, 60 cm. **b**, Linkage between *GA20ox-2* and *sd1* in the rice genome. Left, chromosomal location of *GA20ox-2*; *GA20ox-2* co-segregates with *RG220* on chromosome 1. Right, map position of *sd1*; *sd1* is tightly linked to *RG220* on chromosome 1 (ref. 10). cM, centimorgans. **c**, Mutation sites of the four *sd1* alleles. The *GA20ox-2* gene consists of three exons and two introns. The mutation in each allele is indicated by either an arrow (single-nucleotide substitution) or a line (internal deletion). **d**, Expression of *GA20ox-2* in different organs. Amplification was by polymerase chain reaction with reverse transcription using first-strand complementary DNA derived from different organs. Products were detected by Southern blot DNA analysis; actin DNA was used as a loading control¹¹. The nucleotide sequence of *GA20ox-2* has been deposited in GenBank under accession number AB077025.

carry the same *sd1* allele), which induces a frameshift that creates a stop codon, and that the other three *sd1* alleles encode proteins with amino-acid substitutions (Jikkoku, Calrose 76 and Reimei strains; Fig. 1c). Introducing the *GA20ox-2* gene from the wild-type plant rescued the semi-dwarf phenotype of *sd1*; furthermore, a recombinant *GA20ox-2* protein catalysed the conversion of GA₅₃ to GA₂₀ (results not shown). We conclude that the wild-type *SD1* gene encodes the biosynthesis enzyme GA20ox.

The rice genome carries at least two GA20ox genes (*GA20ox-1* and *GA20ox-2*). *SD1* corresponds to *GA20ox-2*, which is strongly expressed in the leaf blade, stem and unopened flower, whereas *GA20ox-1* is predominantly expressed in the unopened flower (Fig. 1d). The increased expression of *GA20ox-2* in the leaf blade and stem of the wild type would be expected to result in a semi-dwarf phenotype in the enzyme-defective *sd1* mutants, which indeed have shorter leaves and stems. Surprisingly, however, flower formation and fertilization are normal in the mutants, although active gibberellins are important for these events. It is likely that the other GA20ox, which is encoded by *GA20ox-1* and is preferentially expressed in the reproductive organs, enables the flowers in *sd1* plants to develop and be fertilized normally, explaining why

plant height is reduced without seed yield being affected.

The wheat green-revolution gene *Rht* (for 'reduced height')⁴ is a gain-of-function allele caused by a mutation in a transcription factor that is associated with the gibberellin signalling pathway. As wheat has a hexaploid genome, it does not contain recessive alleles such as *sd1* in rice that might otherwise be used to produce a semi-dwarf strain of wheat. Although the genetic and biochemical functions of the rice *SD1* and wheat *RHT* proteins are completely different (that is, recessive versus dominant, loss-of-function versus gain-of-function events, enzyme versus transcription factor, respectively), the products of both genes are linked with gibberellin malfunction. Consequently, manipulation of the biosynthesis or signalling pathways of this growth hormone may offer a means of regulating the height of other important crop plants.

A. Sasaki*, M. Ashikari*, M. Ueguchi-Tanaka*, H. Itoh*, A. Nishimura†, D. Swapan‡, K. Ishiyama§, T. Saito¶, M. Kobayashi§, G. S. Khush‡, H. Kitano||, M. Matsuoaka*

*Bioscience Center and †Graduate School of Bioagricultural Science, Nagoya University, Nagoya 464-8601, Japan
e-mail: makoto@nuagr1.agr.nagoya-u.ac.jp

†Honda R&D, Wako Research Center,
Wako 351-0193, Japan

‡International Rice Research Institute, Manila,
DAPO Box 7777, Philippines

§BioResources Center and ¶Plant Molecular Biology
Laboratory, Riken, Tsukuba 305-0074, Japan

1. Hargrove, T. R. & Cabanilla, V. L. *BioScience* **29**, 731–735 (1979).
2. Dalrymple, D. G. *Development and Spread of High-yielding Rice Varieties in Developing Countries* (Bur. Sci. Tech., Agency Int. Dev., Washington DC, 1986).

3. Khush, G. S. *Genome* **42**, 646–655 (1999).
4. Peng, J. et al. *Nature* **400**, 256–261 (1999).
5. Hooley, R. *Plant Mol. Biol.* **26**, 1529–1555 (1994).
6. Ross, J. J., Murfet, I. C. & Reid, J. B. *Physiol. Plant.* **100**, 550–560 (1997).
7. Hedden, P. & Phillips, A. L. *Trends Plant Sci.* **12**, 523–530 (2000).
8. Toyomasu, T. et al. *Physiol. Plant.* **99**, 111–118 (1997).
9. Xu, Y.-L. et al. *Proc. Natl Acad. Sci. USA* **92**, 6640–6644 (1995).
10. Maeda, H. et al. *Breed. Sci.* **47**, 317–320 (1997).
11. Zhang, W., McElroy, D. & Wu, R. *Plant Cell* **3**, 1155–1165 (1991).

Competing financial interests: declared none.

Materials chemistry

Macroporous crystalline vanadium oxide foam

Porous inorganic solids can be synthesized in complex forms^{1,2}, and here we describe a simple method to create an ultralight, macroporous, crystalline vanadium oxide foam by bubbling oxygen gas produced *in situ* through a viscous vanadium oxide gel. This foaming process could be extended to other metal oxides.

Vanadium oxide (V_2O_5) powder (1 g, or 5.5×10^{-3} mol) is mixed with a solution of a long-chain alkylamine (1-hexadecylamine, $C_{16}H_{33}NH_2$; 1.99 g, or 8.24×10^{-3} mol) in acetone (3 ml). A pasty material is formed, to which is added an aqueous solution of hydrogen peroxide (50 ml, 30%). As H_2O_2 decomposes, oxygen gas evolves and a voluminous bright yellow foam forms spontaneously (Fig. 1). The volume of the foam increases rapidly and reaches 2.5 litres after about 30 min. This new material is ultralight, as only 1 g V_2O_5 is needed to generate the large volume of foam.

Scanning electron microscopy of the foam (Fig. 2a) reveals that it is highly porous and has an irregular, honeycomb-like morphology. Large pores a few micrometres in diameter are surrounded by vanadium oxide walls about 1 μm thick. This unusual pattern resembles the hexagonal cellular structures found in the skeletons of some diatoms and radiolarians³. The mechanical properties of this porous material are poor — the ultralight foam is brittle and can easily be crushed into a powder.



Figure 1 Light fantastic: up to 3 litres of a low-density, bright yellow vanadium oxide foam are formed when 1 g of V_2O_5 reacts with hydrogen peroxide in the presence of hexadecylamine.

The vanadium oxide walls are crystalline, and X-ray diffraction reveals five sharp peaks that can be assigned to the 001 peaks of a lamellar solid with a preferred orientation (Fig. 2b). The first 001 peak gives a basal distance, d , of 33.4 Å, which is much larger than the distance between vanadium oxide planes in the crystalline oxide or in $V_2O_5 \cdot 1.6H_2O$ gels ($d = 11.5$ Å). This basal distance corresponds to that noted for hexadecylamine molecules intercalated in vanadium oxide gels in which the alkyl chains are perpendicular to the oxide planes⁴.

Contrary to observations of hybrid organic–inorganic vanadium oxide com-

pounds, the bright yellow coloration of our vanadium oxide foam indicates that all the vanadium ions are probably in their highest oxidation state (V^{5+}). The hybrid foams contain no water, even when dried at room temperature, and they have amphiphilic alkylamines intercalated between vanadium oxide layers and are highly hydrophobic, separating readily at the surface of an aqueous solution.

Vanadium oxide reacts with hydrogen peroxide to give $V_2O_5 \cdot nH_2O$ gels⁵ that have a layered structure composed of negatively charged vanadium oxide ribbons. These gels can intercalate a wide variety of inorganic and organic species, such as protonated long-chain alkylamines⁶. In the foaming process described here, gelation and intercalation both occur when the 1-hexadecylamine and hydrogen peroxide solutions are added to vanadium oxide powder. The intercalation of hexadecylamine between the V_2O_5 layers creates a pasty solid, while the hydrogen peroxide decomposes spontaneously in the presence of vanadium oxide. Large pores are formed by oxygen gas released through the viscous gel in the presence of surfactants: as they escape, the oxygen bubbles cause the formation of a voluminous polygonal foam, without destroying the crystalline structure of the oxide network of surfactant molecules intercalated between the layers of vanadium oxide.

Meso- and macroporous solids are usually made in the presence of organic templates which then have to be removed without damaging the inorganic shell^{7,8}. In the foaming method we describe here, amines are simply intercalated between vanadium oxide planes and do not act as templates for the formation of macropores. The experimental requirements are not crucial — foams can form with amine/ V_2O_5 molar ratios ranging from 0.5 to 2.5, other surfactants can be used and the process can be extended to many different oxides (for example, those of molybdenum (MoO_3), titanium (TiO_2) and zirconium (ZrO_2)).

G. T. Chandrappa*†, Nathalie Steunou*, Jacques Livage*

*Chimie de la Matière Condensée, Université Paris VI, 4 place Jussieu, 75252 Paris, France
e-mail: livage@ccr.jussieu.fr

†Present address: Department of Chemistry, Bangalore University, 560001 Bangalore, India

1. Mann, S. *Angew. Chem. Int. Edn* **39**, 3392–3406 (2000).
2. Mann, S. & Ozin, G. A. *Nature* **382**, 313–318 (1996).
3. Williams, R. J. P. in *Biomineralization* (eds Mann, S., Webb, J. & Williams, R. J. P.) 1–34 (VCH, New York, 1989).
4. Livage, J. *Chem. Mater.* **3**, 578–593 (1991).
5. Alonso, B. & Livage, J. *J. Solid State Chem.* **148**, 16–19 (1999).
6. Liu, P., Moudrakovski, I. L., Liu, J. & Sayari, A. *Chem. Mater.* **9**, 2513–2520 (1997).
7. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
8. Yang, P., Zhao, D., Margolese, D. I., Chmelka, B. F. & Stucky, G. D. *Nature* **396**, 152–155 (1998).

Competing financial interests: declared none.

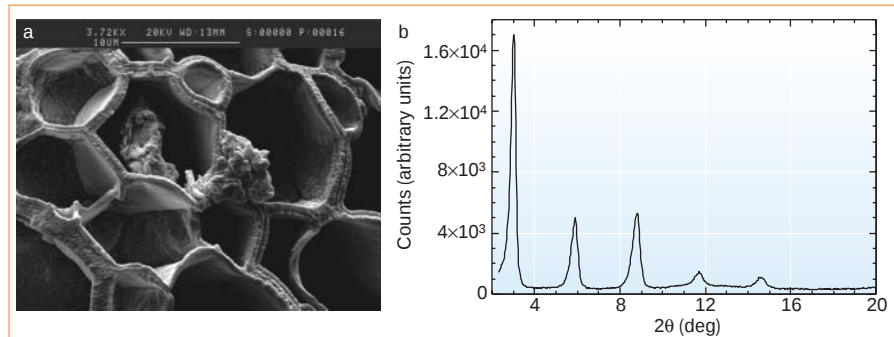


Figure 2 Structure of vanadium oxide foam. **a**, Scanning electron micrograph of vanadium oxide foam, showing its honeycomb-like structure. **b**, The vanadium oxide walls are crystalline, as indicated by the 001 peaks of the X-ray diffraction diagram.

Structure of the Cul1–Rbx1–Skp1–F box^{Skp2} SCF ubiquitin ligase complex

Ning Zheng^{*†}, Brenda A. Schulman^{*‡}, Langzhou Song^{*†}, Julie J. Miller^{*†}, Philip D. Jeffrey^{*}, Ping Wang^{*†}, Claire Chu[§], Deanna M. Koepp[§], Stephen J. Elledge^{‡§}, Michele Pagano^{||}, Ronald C. Conaway^{||}, Joan W. Conaway^{||}, J. Wade Harper[§] & Nikola P. Pavletich^{*†}

^{*} Cellular Biochemistry and Biophysics Program, and [†] Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York 10021, USA
[§] Verna and Marrs McLean Department of Biochemistry and Molecular Biology, and Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

^{||} Department of Pathology and Kaplan Comprehensive Cancer Center, NYU Medical Center, New York 10016, USA

[‡] Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA

SCF complexes are the largest family of E3 ubiquitin–protein ligases and mediate the ubiquitination of diverse regulatory and signalling proteins. Here we present the crystal structure of the Cul1–Rbx1–Skp1–F box^{Skp2} SCF complex, which shows that Cul1 is an elongated protein that consists of a long stalk and a globular domain. The globular domain binds the RING finger protein Rbx1 through an intermolecular β -sheet, forming a two-subunit catalytic core that recruits the ubiquitin-conjugating enzyme. The long stalk, which consists of three repeats of a novel five-helix motif, binds the Skp1–F box^{Skp2} protein substrate-recognition complex at its tip. Cul1 serves as a rigid scaffold that organizes the Skp1–F box^{Skp2} and Rbx1 subunits, holding them over 100 Å apart. The structure suggests that Cul1 may contribute to catalysis through the positioning of the substrate and the ubiquitin-conjugating enzyme, and this model is supported by Cul1 mutations designed to eliminate the rigidity of the scaffold.

Ubiquitin-dependent proteolysis regulates protein abundance and serves a central regulatory function in many biological processes (reviewed in ref. 1). The ubiquitination of a target protein is mediated by the ubiquitin–protein ligases, which represent a diverse super-family of proteins and protein complexes. The SCF (Skp1–Cullin–F-box protein) and SCF-like complexes are the largest family of ubiquitin–protein ligases and ubiquitinate a broad range of proteins involved in cell cycle progression, signal transduction and transcription^{2,3}. Deregulation of SCF-dependent proteolysis can contribute to neoplastic transformation, with the loss of Cyclin E and β -catenin ubiquitination and the accelerated proteolysis of p27^{Kip1} being frequent events in cancer (refs 4–8). In addition, the Fbw7/hCdc4/hAgo SCF subunit is mutated in a number of breast- and ovarian-cancer cell lines^{4,6}, the Skp2 SCF subunit can cooperate with Ras to transform cells^{9,10}, and the von Hippel–Lindau tumour suppressor is a subunit of an SCF-like E3 (refs. 11,12).

Ubiquitin-protein ligases (also known as E3s) act at the last step of a three-enzyme cascade involving the ubiquitin-activating (E1) and ubiquitin-conjugating (E2) enzymes¹³. The E3 mediates the transfer of ubiquitin from the E2 to the substrate protein by promoting the formation of an isopeptide bond between the Ub carboxy-terminus and specific lysine side chains on the substrate. E3s bind both the protein target and a cognate E2 and have a central role in conferring specificity to the ubiquitination pathway. Two functionally distinct types of E3s have been identified to date. HECT-type E3s catalyse ubiquitination by first forming an E3–ubiquitin thioester intermediate. RING-type E3s do not appear to form such an intermediate. They are characterized by the presence of a RING zinc finger domain that binds the E2, but otherwise have diverse subunit compositions and sequences¹⁴. The mechanism by which they promote ubiquitination has not been well understood.

The SCF complexes are RING-type E3s that consist of Cul1 (776 residues), Rbx1 (108 residues), Skp1 (163 residues) and a member of the F-box protein family (~430 to >1,000 residues). Rbx1, which contains the RING domain, and Cul1 form a catalytic core complex that recruits a cognate E2, the variable F-box protein subunit binds the substrate^{15–21}, and Skp1 serves as an adapter that links the F-box

protein to Cul1 (ref. 16). F-box proteins are characterized by an amino-terminal 40-residue F-box motif that binds Skp1 (ref. 15), followed by protein–protein interaction modules such as leucine-rich repeats or WD-40 repeats that bind substrate^{15,16}. The large number of F-box proteins in eukaryotic genomes (at least 38 in human²²) is thought to allow for the specific ubiquitination of a large number of functionally and structurally diverse substrates. Human SCF complexes with demonstrated E3 activity include the SCF^{Skp2} complex (the superscript denotes the F-box protein), which ubiquitinates the Cdk-inhibitor p27^{Kip1}, the SCF ^{β -TrCP}, which ubiquitinates β -catenin and I κ B, and the SCF^{Fbw7/hCdc4/hAgo}, which ubiquitinates cyclinE (refs 2, 4–6). The ubiquitination of p27^{Kip1} but not of other known SCF substrates also requires the Cks1 protein^{23,24}.

In addition to multiple F-box proteins, most higher eukaryotes also contain multiple homologues of the other SCF subunits, including two Rbx1 and five cullin family members (paralogues) conserved from *C. elegans* to humans^{19–21,25}. Most of these homologues remain poorly understood, except that Cul2 has been shown to have a Cul1-like function in assembling a ubiquitin-ligase complex that includes Rbx1, ElonginC and the SOCS-box family protein VHL (reviewed in ref. 26). Homology to the Cul1 and Rbx1 SCF subunits has also been found in the 11-subunit Anaphase-Promoting Complex (APC) E3 that promotes the ubiquitination of several mitotic regulatory proteins^{27,28}.

Here we report the 3.2-Å structure of the quaternary complex containing Cul1, Rbx1, Skp1 and the F-box of Skp2, and the 3.0-Å structure of the Cul1–Rbx1 complex. These structures, in conjunction with the published structure of the Skp1–Skp2 complex²⁹ and mutational analyses, provide a framework for understanding the SCF-mediated ubiquitination reaction, the ability of SCFs to target a large number of diverse substrates, and the roles of cullin and Rbx1-like domains in other proteins.

Overall structure of the SCF

The Cul1–Rbx1–Skp1–F box^{Skp2} complex has a highly elongated structure with Rbx1 and the Skp1–F box^{Skp2} complex segregated to opposite ends (Fig. 1). This arrangement is organized by Cul1, which interacts with all three subunits and serves as a scaffold. Cul1 consists of a 415-amino-acid N-terminal helical region (hereafter

[‡] Present address: Departments of Structural Biology and Tumor Cell Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

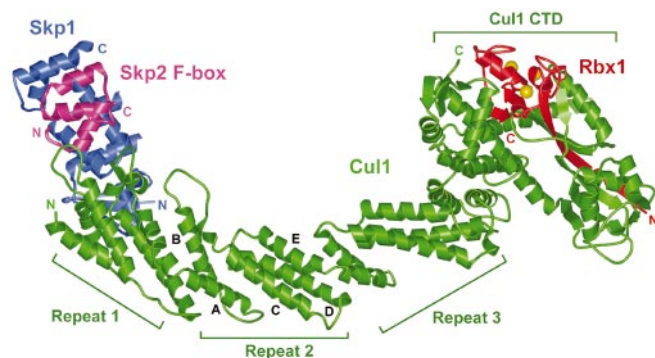


Figure 1 Overall structure of the Cul1–Rbx1–Skp1–F box^{Skp2} quaternary complex. Cul1, Rbx1, Skp1 and the F-box of Skp2 are coloured in green, red, blue and magenta, respectively. The five helices that make up the cullin-repeat motif are labelled for the second repeat. Figures were prepared with the programs MOLSCRIPT⁴⁹, GL_RENDER and POVray (L. Essar, personal communication).

NTD) that adopts a long stalk-like structure and binds the Skp1–F box^{Skp2} complex, and a 360-amino-acid C-terminal globular α/β domain (hereafter CTD) that binds Rbx1 (Fig. 1).

The Cul1 NTD stalk consists of three repeats of a novel five-helix structural motif (the cullin repeat). The three cullin repeats are arranged in a regularly repeating fashion resulting in an arc-shaped structure spanning ~ 110 Å (Fig. 1). The N-terminal tip of the first repeat binds the Skp1–F box^{Skp2} complex. At the opposite end, the C-terminal two helices of the third repeat pack with a four-helix bundle from the Cul1 CTD.

The Cul1 CTD contains several structural units that assemble

around the Rbx1 protein to form a single globular unit. Part of the Cul1 CTD forms a five-stranded intermolecular β -sheet that contains as its second β -strand an N-terminal sequence of Rbx1. The rest of the Cul1 CTD forms a ~ 30 -Å-wide groove wherein the Rbx1 RING domain is embedded. The RING domain of Rbx1 is a variant of the canonical RING motif, with a 20-amino-acid insertion that forms a third zinc coordination site.

Structure of Cul1

Cullin repeats of the N-terminal domain

The cullin-repeat motif consists of two short helices (A and B) and three long helices (C to E; Fig. 2a). The C, D and E helices are arranged in a three-helix bundle, and the A and B helices pack with the N-terminal half of the bundle. This ~ 120 -amino-acid cullin-repeat motif is distinct from known helical repeats such as HEAT and TPR (reviewed in ref. 30).

The three cullin repeats have similar overall structures but the precise arrangement of their helices varies by up to 2.5 Å. Repeats 2 and 3 are the most similar pair, and can be superimposed with a root mean square deviation (r.m.s.d.) of 1.9 Å for 99 C α atoms. Repeat 1 is more divergent owing to shifts and rotations in helices A and B, which function as an N-terminal cap, and also because of a two-helix insertion (H5 and H6, Fig. 2a and Supplementary Information). Both of these unique features of repeat 1 are associated with Skp1 binding, as discussed later. There is no clear amino-acid sequence motif associated with the three repeats, although they do share a similar pattern of hydrophobic residues.

Adjacent repeats pack through helices D and E of one repeat interacting with helices A and B of the next repeat (Fig. 1). Helix E of one repeat and helix A of the next repeat belong to a nearly continuous long helix (Fig. 1). Adjacent repeats are related by a

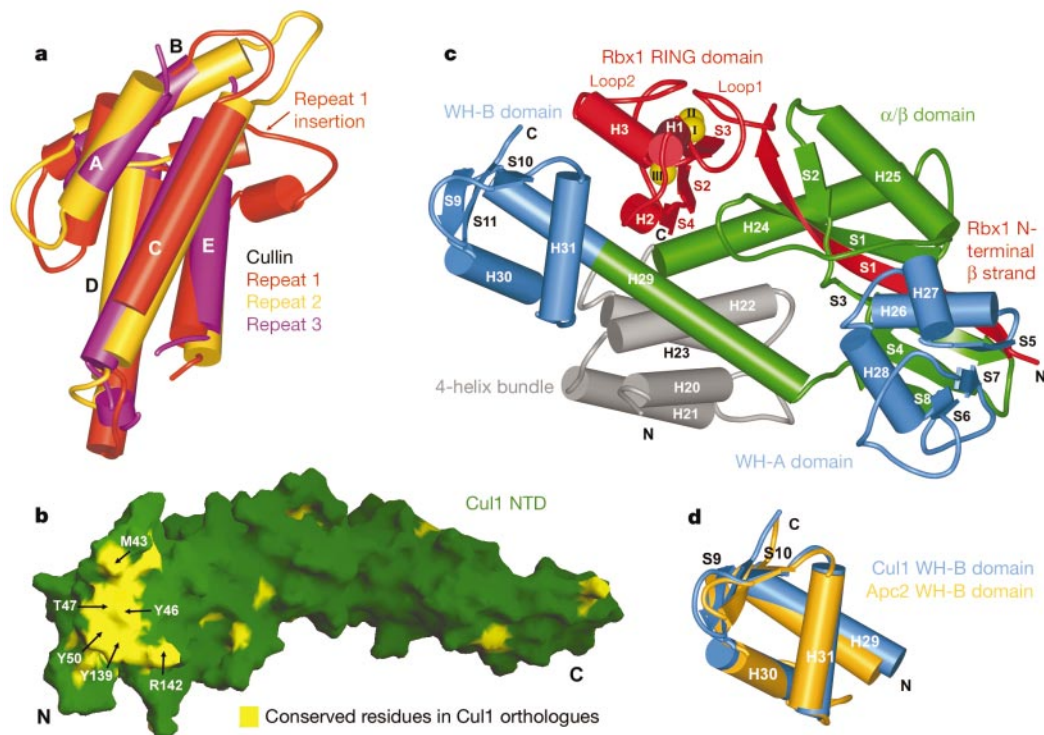


Figure 2 Structural elements of the Cul1 N- and C-terminal domains (NTD and CTD). **a**, Superposition of the first, second and third cullin repeats, coloured in red, orange and magenta, respectively. The five helices of the cullin-repeat-motif are labelled A to E, and the two-helix insertion of repeat 1 is marked. **b**, Surface representation of the first cullin repeat showing the patch of surface residues invariant in Cul1 orthologues from yeast, fly, worm, mouse and human. The patch is coloured yellow and the underlying residues are

labelled. **c**, Overview of the Cul1 CTD with Rbx1 bound. The 4HB, α/β , WH-A and WH-B domains are coloured in grey, green, blue, and blue, respectively. Rbx1 is in red and the three zinc ions are in yellow. **d**, Superposition of the WH-B domains of yeast Apc2 (orange) and of Cul1 (cyan). The orientation of the Cul1 WH-B domain is as in Fig. 2c. The secondary structure elements of Cul1 WH-B are labelled.

~37 Å translation and a ~27° rotation and this inter-repeat relationship is remarkably conserved. Applying the repeat 1 → 2 transformation twice to repeat 1 positions it within 2.5 Å of repeat 3.

The conservation in the NTD sequence among Cul1 orthologues (see Supplementary Information) maps either to residues buried in the hydrophobic core, or to residues that cluster together on a surface patch at the NTD N-terminal tip that is the Skp1–F box^{Skp2} binding site (Fig. 2b). The lack of other conserved surface patches suggest that the main function of the NTD is to serve as a long spacer with a binding site for Skp1–F box^{Skp2} at one end.

C-terminal domain

The CTD structure consists of a four-helix bundle (4HB), an α/β domain with an intermolecular five-stranded β-sheet, which contains as its second strand an N-terminal segment of Rbx1, and two copies of the winged-helix motif (WH-A and WH-B; Fig. 2c). The 4HB connects the CTD to the NTD and also organizes the CTD structure. It packs with the long H24 and H29 helices, which in turn pack with the α/β domain and the two winged-helix domains (Fig. 2c). This results in a V-shaped CTD groove where the RING domain of Rbx1 binds. The importance of the 4HB in organizing the CTD structure is apparent in its high conservation not only among Cul1 orthologues, but also among all Cul1 paralogues. In addition, the yeast *cdc53-1* temperature-sensitive mutation³¹ disrupts a buried salt bridge within the 4HB (between Arg 472 and Glu 485 in human Cul1).

Cullin-homology region

A ~200-amino-acid Cul1 region has been reported to have homology to several other proteins such as the Apc2 subunit of the APC E3, and it has been termed the cullin homology region^{27,28,32}. The cullin homology region starts in the middle of the 4HB and contains the H24 helix and parts of the β-sheet and WH-A domains. The structure suggests that the proteins that have the cullin-homology sequence will also have the rest of the 360-amino-acid CTD structure, even though they typically have no significant sequence homology outside the cullin-homology region. The 4HB, the α/β domain and the WH-A are likely to be present as intact structural units, and their packing interactions with the WH-B would require its presence as well. To explore this further, we crystallized and determined the 2.0-Å structure of the C-terminal 78 amino acids of yeast Apc2, which is a region well beyond the cullin-homology region and has no detectable homology to Cul1 (see Supplementary Information for an Apc2–Cul1 alignment and structure determination of Apc2 WH-B). Figure 2d shows that the Apc2 C-terminus adopts the winged-helix fold and can be superimposed on the WH-B of Cul1 with an r.m.s.d. of 1.66 Å for 65 out of 78 residues.

Rbx1 has a variant RING finger

The Rbx1 structure consists of a 16-residue β-strand (S1) that gets incorporated into the Cul1 CTD β-sheet, followed by a variant RING finger domain (Fig. 2c). The 70-residue Rbx1 RING domain adopts the structure of the canonical RING motif stabilized by two zinc ions (sites I and II, Fig. 3a), but it also has a 20-residue insertion. The insertion contains three additional zinc ligands (Cys 53, Cys 56, and Cys 68), which, together with a fourth zinc ligand from the RING motif (Cys 82), form a new zinc-binding site (site III; Fig. 3a and Supplementary Information). Although the site-III zinc ligands are invariant in Rbx1 orthologues and family members, the rest of the insertion sequence is not.

Comparison of the Rbx1 RING domain structure with that of the c-Cbl RING E3 bound to an E2 (ref. 33) reveals that the hydrophobic groove the c-Cbl RING domain uses to bind the E2 is present in Rbx1 as well (Fig. 3a). The two loops and helix that make up the c-Cbl groove have a conserved structure in Rbx1 (Loop1, Loop2 and H3 C-terminus; Fig. 3a), and the hydrophobic c-Cbl residues that

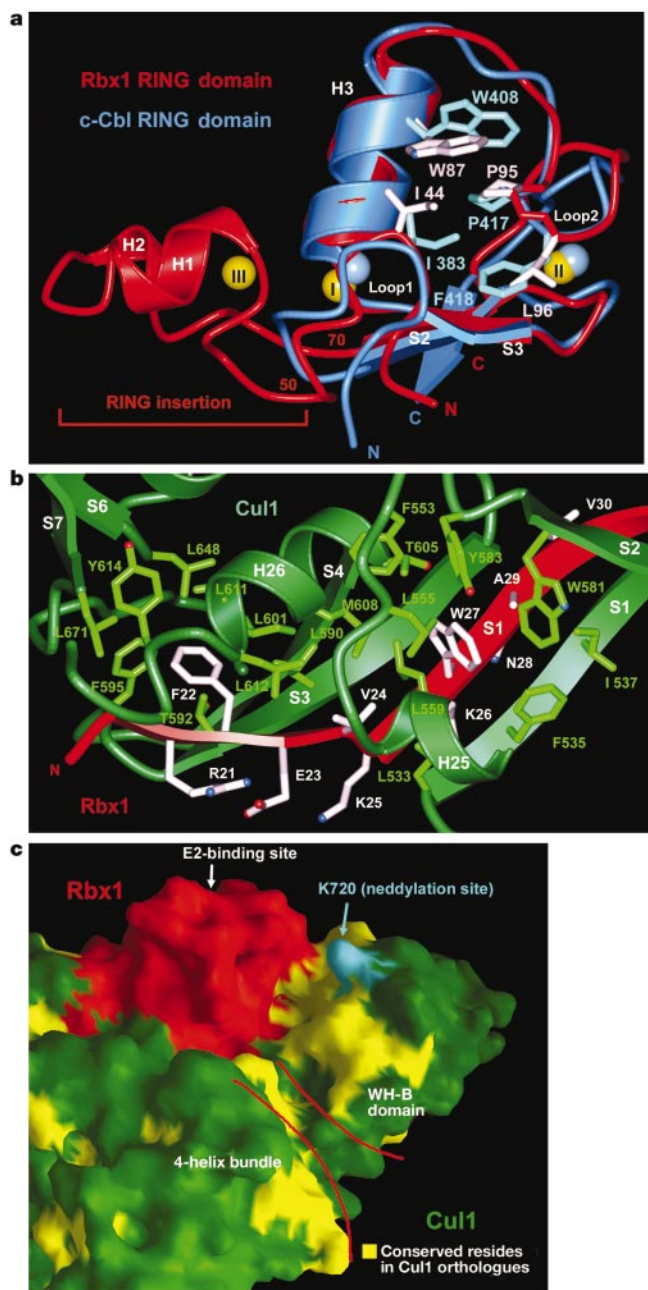


Figure 3 Interactions between Rbx1 and the Cul1 CTD. **a**, The Rbx1 RING domain has a core part, consisting of two large loops (Loop1 and Loop2), an α-helix (H3), and a three-stranded β-sheet (S2, S3 and S4), that is very similar to the RING domain of the c-Cbl E3. It also has an insertion (residues 50–70) that is stabilized by a third zinc ion (site III). Rbx1 and c-Cbl side chains are coloured pink and cyan, respectively. **b**, Close-up view of the intermolecular β-sheet formed by Rbx1 (red) and Cul1 (dark green). The Cul1 residues that contact Rbx1 are shown in light green, and the Rbx1 residues in pink. **c**, Surface representation of the complex in the vicinity of the neddylation site (Lys 720) of Cul1. The surface of Rbx1 and Cul1 is in red and green, respectively, except for the residues conserved in Cul1 orthologues, which are in yellow. The surface of the exposed side chain of Lys 720 is coloured cyan. Solid red lines delineate the approximate boundary between the WH-B and 4HB domains. The E2-binding site, proposed according to c-Cbl homology and mutagenesis analyses, is indicated. The figure was made with GRASP⁵⁰.

contact the E2 either are conserved or maintain their hydrophobic character in Rbx1 (Ile 44, Trp 87, Pro 95 and Leu 96; Fig. 3a), suggesting that Rbx1 would bind its cognate E2s in a similar manner. This model is supported by the observations that mutation of any one of the zinc ligands at sites I and II, which have a central

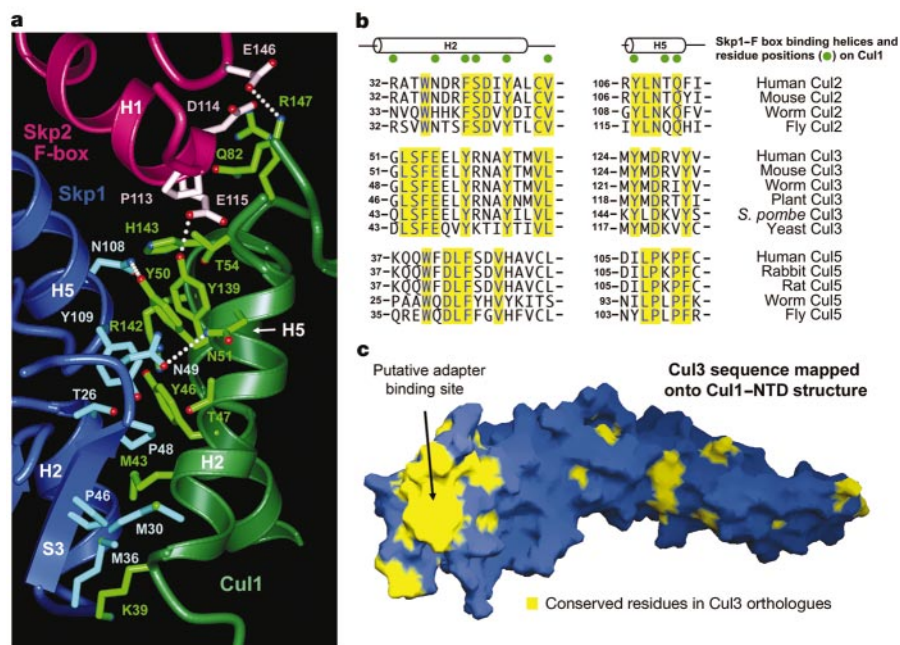


Figure 4 Cul1-Skp1-F box^{Skp2} interface and the putative protein-binding site on other cullins. **a**, View of the interface between Cul1 (dark green) and Skp1-F box^{Skp2} (blue and magenta) in an orientation rotated 90° about the vertical axis of Fig. 1. The residues of Cul1, Skp1, and F-box^{Skp2} that are involved in the interactions are shown in light green, cyan and pink colours, respectively. Hydrogen bonds are indicated by white dotted lines. **b**, Sequence alignments of the orthologues of Cul2, Cul3 and Cul5 in the regions that correspond to the Skp1-F box^{Skp2} binding site of Cul1. Invariant residues are highlighted

in yellow. Cul4B is not shown because only human and fly sequences are available. The published Cul4A sequence lacks most of the first cullin repeat. **c**, The sequence of Cul3 was mapped onto the structure of the Cul1 NTD on the basis of sequence homology, and the resulting model's surface was coloured according to conservation in Cul3 orthologues. The model surface shows that the residues identical in Cul3 orthologues (yellow) cluster on the same surface region as the Skp1-F box^{Skp2} binding site of Cul1 (Fig. 2b). A Cul5 model constructed the same way shows a similar feature.

role in stabilizing the hydrophobic groove structure, abolished the Rbx1-Cul1 E3 ligase activity, but mutation of the zinc ligands at site III, which is not near the putative E2-binding site, did not appreciably affect Rbx1's function^{19,20,34}. To further support the proposed E2-binding site, we constructed the Trp 87 → Ala, Lys 89 → Ala, Thr 90 → Ala and Arg 91 → Ala mutants of solvent-exposed residues at the C-terminus of the H3 helix, and found that they are defective in complementing an *rbx1Δ* yeast strain (see Supplementary Information).

Cul1-Rbx1 binding

The Cul1-Rbx1 complex buries 3,396 Å² of surface area, with 1,734 Å² contributed by the Rbx1 S1 β strand and 1,662 Å² by the RING domain. The 16-residue Rbx1 β strand, which contains a kink in the middle, makes β-sheet backbone hydrogen bonds with the S1, S2 and S3 strands of Cul1, and it participates in the formation of the α/β hydrophobic core. The conserved Phe 22 and Trp 27 of Rbx1 have a central role, making multiple van der Waals contacts to side chains from the WH-A, H25, S1, S2 and S3 of Cul1 (Fig. 3b). The Cul1 residues involved in these interactions are highly conserved in Cul1 orthologues and paralogues (Fig. 3b and see Supplementary Information), indicating that other combinations of cullin and Rbx family members will form a similar intermolecular β-sheet. After the Rbx1 β strand, a tryptophan-rich Rbx1 region (Trp 33-Ala 34-Trp 35) packs both with the α/β domain and the following RING domain, resulting in a continuous surface between the two.

The Rbx1 RING domain packs in a V-shaped groove formed by the Cul1 α/β domain on one side, and the WH-B domain on the other side (see Supplementary Information). The Rbx1 RING-Cul1 interactions are not extensive and the residues involved are not as conserved as those in the Rbx1 strand-Cul1 α/β interface.

The intermolecular β-sheet appears to be the primary mechanism of Rbx1 recruitment, and this would explain the observation that mutations in the zinc-chelating residues at sites I and II did not

disrupt the association between Rbx1 and Cul1, even though they abolished SCF function^{20,34}.

The Rbx1 S1 strand is conserved in Apc11, which is the RING finger subunit of the APC E3. The Phe 22 and Trp 27 Rbx1 residues that pack in the Cul1 α/β hydrophobic core correspond to Val 3 and Trp 8 in Apc11, and the Trp 33-Ala 34-Trp 35 sequence that bridges the S1 strand to the RING domain, corresponds to Trp 14-Leu 15-Trp 16 (see Supplementary Information for an Apc11-Rbx1 alignment). Coupled to the cullin-homology-region conservation between Cul1 and Apc2, these observations indicate that Apc11 binds to Apc2 through a similar strand-insertion process.

Nedd8 modification of Cul1

Cul1 is modified through the covalent attachment of the ubiquitin-like small molecule Nedd8 (reviewed in ref. 2). *In vitro*, neddylation enhances the activities of the SCF^{β-TrCP} and SCF^{Skp2} towards the IκB and p27^{Kip1} substrates, respectively^{35,36}. Cul1 is neddylated at Lys 720, which is on the WH-B domain and is positioned at the rim of a cleft formed by conserved residues from the WH-B, 4HB and Rbx1 RING domain (Fig. 3c). The proximity of Lys 720 to the Rbx1 RING domain (11 Å to the nearest Rbx1 side chain) supports the model that Nedd8 might modulate the binding and positioning of the E2 or the E2-ubiquitin conjugate^{37,38}.

Cul1-Skp1-F box^{Skp2} ternary interface

The structures of a nearly full-length functional Skp1-Skp2 and of the truncated Skp1-F box^{Skp2} complexes have been described²⁹, and the Skp1-F box^{Skp2} structure is essentially unchanged in the quaternary complex (Fig. 1). Cul1 binds the Skp1-F box^{Skp2} complex through the N-terminal portion of its first cullin repeat interacting with both Skp1 and the F-box motif (Fig. 1, 4a). A Cul1 hydrophobic surface consisting of H5, which corresponds to an insertion in the cullin-repeat motif, and H2 packs with hydrophobic and polar residues from H2, S3 and H5 of the BTB/POZ fold of Skp1

and from the N-terminal portions of the H1 and H3 helices of the Skp2 F-box (ref. 29), burying a total of 1,780 Å² area.

The centre of the interface contains residues highly conserved in Cul1 and Skp1 orthologues (Met 43, Tyr 46, Thr 47, Tyr 50, Tyr 139 and Arg 142 of Cul1 and Asn 49, Asn 108 and Tyr 109 of Skp1; Fig. 4a). Mutation of these residues either abolishes complex formation or significantly reduces the affinity of the complex *in vitro* (see Supplementary Information). The adjacent Cul1–F box^{Skp2} interactions involve two Skp2 residues conserved among most but not all F-box proteins (Pro 113 and Glu 115 from the H1 helix; Fig. 4a and ref. 29). Superposition of the full length Skp1–Skp2 complex with the Skp1–F box^{Skp2} complex shows that the rest of Skp2 would not be within contact distance of Cul1.

Putative adapter binding sites in other cullins

Surface residues in the Skp1–F box^{Skp2} binding site of Cul1 are conserved only in Cul1 orthologues but not in paralogues. However, when orthologues of individual cullins (2, 3, 4B and 5) are compared, their putative H2 and H5 helices emerge as the best-conserved portion of their NTDs, suggesting that they too have a protein-binding site in their first cullin repeat (Fig. 4b, c). Cul2 has already been shown to bind the Skp1-like ElonginC adapter protein, and the structure suggests their association is similar to that of Cul1–Skp1. In fact, several Cul1–Skp1 interface residues (Tyr 50/ Tyr 139 of Cul1; Thr 42/Met 45/Asn 108 of Skp1) are conserved in Cul2 and ElonginC (Tyr 43/Tyr 107 of Cul2; Thr 25/Met 29/Asn 92 of ElonginC). However, the more central interface residues in the Cul1–Skp1 complex (Tyr 46 of Cul1 and Tyr 109 of Skp1) differ in Cul2 and ElonginC, suggesting that they contribute to specificity (Fig. 4b).

Rigidity of Cul1 scaffold required for SCF function

The lack of flexible linkages in the Cul1–Rbx1–Skp1–F box^{Skp2} is reminiscent of the rigidity observed in the Skp1–Skp2 (ref. 29) complex and the c-Cbl E3 (ref. 33). In c-Cbl, the substrate-binding SH2 and E2-binding RING domains are tightly coupled, with inactivating mutations mapping to the structured linker in between the two³³. Similarly, the NTD and CTD domains of Cul1 are rigidly linked, and the temperature-sensitive *cdc53-2* mutation in yeast Cul1 maps to a structural residue (Gly 332 of human Cul1) in the third cullin repeat that bridges the NTD and CTD (ref. 31). To start investigating the importance of the rigid architecture of the Cul1 scaffold, we sought to construct a Cul1 mutant where the NTD and CTD interface is disrupted, and where the two domains are linked by a flexible linker (Fig. 5a). The flexibly linked Cul1 mutant is able to bind the phosphorylated p27 substrate in a Skp1–Skp2–Cks1 dependent manner (Fig. 5b) and also to promote the substrate independent polymerization of ubiquitin (see Supplementary Information); however, it fails to ubiquitinate p27^{Kip1} *in vitro* (Fig. 5c). These results suggest that the rigidity of the Cul1 scaffold and of the entire SCF is important for the E3 activity, although we cannot conclusively rule out the possibility that other factors may contribute to the loss of function of the flexible Cul1 (see Supplementary Information).

Model of the SCF^{Skp2}–E2 complex

We have built a model of the intact SCF^{Skp2}–E2 complex by superimposing Cul1–Rbx1–Skp1–F box^{Skp2} on Skp1–Skp2 (ref. 29), and by docking the UbC^{H7} E2 onto the Rbx1 RING domain based on the c-Cbl–UbC^{H7} structure (Fig. 6). In this model, the Skp2 leucine-rich repeat domain and the E2 are on the same side of the SCF complex. The tip of the leucine-rich repeat domain points toward the active site cysteine of the E2, with a distance of ~50 Å in between them. This distance could readily be bridged by the portion of p27 between phosphothreonine Thr 187, required for Skp1–Skp2–Cks1 binding, and the candidate ubiquitination sites at lysines 134, 153, and 165 (ref. 39). This architecture

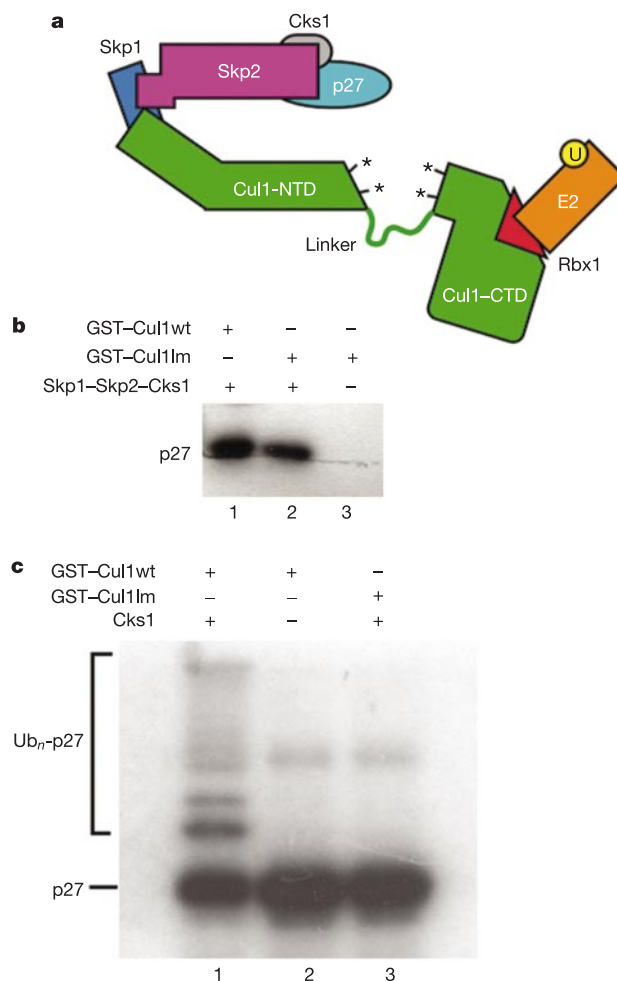


Figure 5 The rigidity of the Cul1 scaffold is important for SCF E3 activity. **a**, The schematic design of the flexible Cul1 mutant. The interface between the third cullin repeat and 4HB was disrupted by hydrophobic-to-polar mutations (Val 367 → Arg, Leu 371 → Asp, Leu 421 → Lys, Val 451 → Asp and Val 452 → Glu), and a flexible linker sequence (KGTRGKGSGPEG) was introduced between the NTD and CTD. **b**, The Cul1 linker mutant retains the ability to bind phosphorylated p27, in a manner dependent on the presence of Skp1, Skp2 and Cks1. **c**, The SCF^{Skp2} complex with the wild-type (WT) Cul1 (lane 1) but not the linker mutant Cul1 (lane 3) promotes the Cks1-dependent polyubiquitination (Ub_n) of p27 in an *in vitro* ubiquitination assay reconstituted with purified components. The high relative-molecular-mass band present in lanes 2 and 3 is due to the phosphorylation of one of the ten reaction components, and is independent of Cks1 (lane 2).

of the SCF suggests that the Cul1 subunit evolved the long stalk to separate the substrate binding and the catalytic activities so that substrates of varying sizes and with varying spacings between their ubiquitinated lysines and their SCF-binding motifs can be accommodated.

Implications for the function of the SCF-like E3s

Other than selecting the protein to be ubiquitinated, the role of the SCF in the ubiquitination reaction has not been clear. The possibility that the SCF may form an E3–ubiquitin intermediate like HECT E3s has previously been ruled out^{21,40}, and the structural data presented here would rule out the model that it may participate in acid/base catalysis, as there are no SCF residues in the proximity of the E2 active site in our SCF–E2 model. On the other hand, the rigidity observed in the Cul1–Rbx1–Skp1–F box^{Skp2} structure, in conjunction with mutational data, suggests that RING E3s may facilitate ubiquitin transfer by positioning the substrate protein in

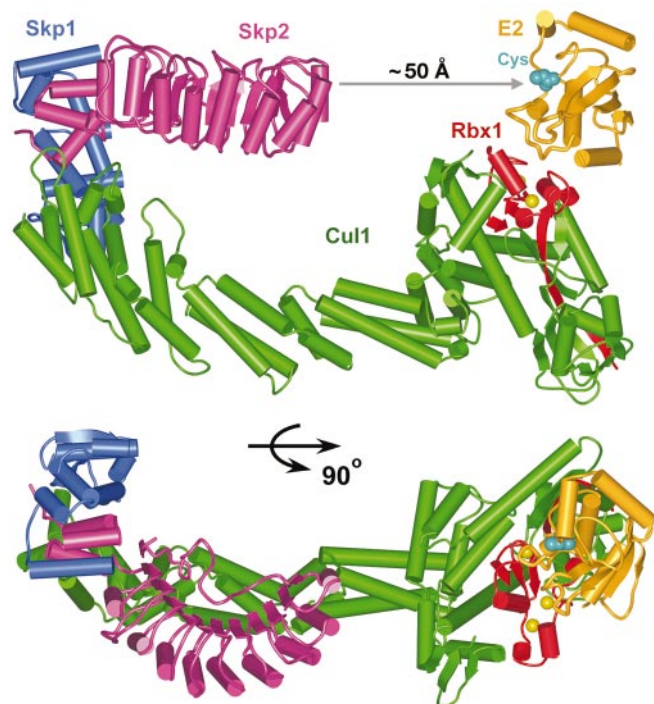


Figure 6 Model of the SCF^{Skp2}-E2 complex. Cul1, Rbx1, Skp1, Skp2, and an E2 are coloured in green, red, blue, magenta, and orange, respectively. The active site cysteine of the E2, where ubiquitin would be covalently attached, is shown in space filling representation and coloured in cyan. The grey arrow indicates the 50-Å gap between the tip of the Skp2 leucine-rich repeat domain and the E2 active site.

an optimal manner, such that the lysine side chains to be ubiquitinated are presented in the immediate vicinity of the E2 active site. With substrates that have a large size or considerable flexibility, this may significantly enhance the reaction rate. The extent to which this positioning is achieved is still unclear, and could range from raising the local concentration of the lysine(s) around the E2 active site, to positioning and orienting the epsilon amino group of the lysine for attack on the E2-ubiquitin thioester bond. A positioning effect may also explain lysine specificity. Among the few ubiquitinated proteins that have been characterized in this much detail, IκB and SnN are ubiquitinated at specific lysine residues by the SCF and SCF-like complexes^{41,42}. Because no sequence or structural motif has been identified for a ubiquitination site so far, spatial restraints imposed by the E3s might be a major determinant for such specificity. □

Methods

Protein expression and purification

The crystallized human Cul1-Rbx1 complex was produced by co-expressing the two full-length proteins in insect cells from separate viruses, with Cul1 expressed as a glutathione S transferase (GST) fusion protein. The complex was purified first by glutathione affinity, and following cleavage of GST with thrombin, by cation exchange and gel filtration chromatography. The Cul1-Rbx1 complex was concentrated to about 10 mg ml⁻¹ by ultrafiltration in 20 mM 2-(N-Morpholino) ethanesulphonic acid, 200 mM NaCl, 5 mM dithiothreitol (DTT), pH 6.5. To assemble the quaternary complex, *E. coli*-produced Cul1-Rbx1 (see Supplementary Information) was mixed with Skp1-F box^{Skp2} complexes purified as described²⁹ at 1:1 molar ratio with a final complex concentration of 10–20 mg ml⁻¹.

Crystallization and data collection

Crystals were grown at 4 °C by the hanging-drop diffusion method. The Cul1-Rbx1 crystals grew from 100 mM Tris-HCl, 12–18% ethanol, 200–500 mM NaCl, 5 mM DTT, pH 8.0. They form in space group P2₁, with *a* = 113.9 Å, *b* = 50.0 Å, *c* = 135.9 Å, β = 107.8° and contain one complex in the asymmetric unit. Crystals of the Cul1-Rbx1-Skp1-F box^{Skp2} complex grew from 100 mM BTP, 7–9% PEG4000, 200–400 mM NaCl, 5 mM DTT, pH 7.0. They form in space group P2₁2₁2₁, with *a* = 219.3 Å, *b* = 50.5 Å, *c* = 158.6 Å and contain one complex in the asymmetric unit. Heavy atom soaks were

performed in crystallization buffer lacking DTT with 0.3 mM KAu(CN)₂ (4 h), 0.25 mM K₂PtCl₄ (4 h), 1 mM uranyl acetate (12 h) and a double soak of 0.25 mM KAu(CN)₂ plus 0.25 mM K₂PtCl₄ (4 h). The Cul1-Rbx1 and Cul1-Rbx1-Skp1-F box^{Skp2} data sets were collected at the F1 (λ = 0.943) and A1 (λ = 0.928) MacCHESS beamlines, respectively, at –170 °C. The Cul1-Rbx1 crystals were flash-frozen in crystallization buffer supplemented with 30% ethylene glycol. The spot profiles of the Cul1-Rbx1-Skp1-F box^{Skp2} crystals were very streaky. Flash-freezing the crystals in a cryoprotectant solution with higher pH (100 mM Tris-HCl, pH 8.0, 9% PEG4000, 300 mM NaCl, and 30% PEG400) improved but did not eliminate the streakiness. We presume the high *R*_{sym} of the Cul1-Rbx1-Skp1-F box^{Skp2} dataset and the high *R* factor of the refined model are caused by problems integrating the streaky spots (see Supplementary Information). Data were processed with DENZO and SCALEPACK⁴³.

Structure determination and refinement

The structure of the Cul1-Rbx1 complex was determined by the multiple isomorphous replacement anomalous scattering (MIRAS) method (see Supplementary Information). Phases were calculated with the program SHARP⁴⁴ to 3.4 Å and had a mean figure of merit of 0.72. They were improved by density modification and solvent flattening with the program CNS SOLVE⁴⁵. The MIRAS maps had continuous electron density for the majority of the Cul1-Rbx1 complex (see Supplementary Information). The model was built with the program O (ref. 46), and was improved by several cycles of manual rebuilding and refinement (simulated annealing, positional and grouped *B* factor) with CNS (ref. 45). The model was confirmed with simulated annealing omit maps and also by collecting anomalous dispersion data at the sulphur *F*^o maximum at the X4A beamline of the NSLS at Brookhaven (λ = 1.77 Å) (data not shown). Anomalous dispersion maps calculated with the MIRAS phases showed clear density for 21 of the 30 cysteine and methionine residues of Cul1 and for the three zinc ions of Rbx1. The refined model contains residues 17–55 and 83–776 of Cul1 and residues 19–106 of Rbx1. 86% of residues are in the most favoured region and 14% in the allowed region of the ramachandran plot. The structure determination of the Apc2 C-terminal WH-B domain is described in the Supplementary Information.

The structure of the Cul1-Rbx1-Skp1-F box^{Skp2} complex was determined by molecular replacement with the program AMORE⁴⁷ using the structures of the Cul1 NTD and the Cul1 CTD bound to Rbx1 as search models. 3.2-Å electron-density maps calculated with phases derived from the rigid-body-refined Cul1-Rbx1 model had clearly interpretable density for the portion of Skp1 that binds Cul1 and also for the F-box, which is involved in crystal packing contacts. Skp1-F box^{Skp2} was built into the difference density using its previously determined 1.8-Å structure, and was improved by several cycles of manual rebuilding and refinement (simulated annealing, positional and grouped *B* factor) with CNS⁴⁵. The refined model contains residues: 15–55, 82–149, 154–216, 225–776 of Cul1; 19–106 of Rbx1; 2–37, 44–68, 84–140 of Skp1; and 109–149 of Skp2, with 83% of residues in the most favoured region and 17% in the additionally allowed region of the ramachandran plot. The relative orientation of the NTD and CTD does not change appreciably in the two crystals, and the two Cul1 structures can be superimposed with an r.m.s.d. of 1.70 Å in the Cα positions of 690 out of 725 ordered residues. The NTD-CTD interface is invariant, with the two interfaces (residues 299–494 corresponding to the third cullin repeat–4HB segment) superimposing with an r.m.s.d. of 0.98 Å.

In vitro ubiquitination and p27 binding assays

The full-length p27 protein, phosphorylated Cdk2-CyclinA complex, and full-length Skp1-Skp2 were prepared as previously described^{29,48}. Cks1 was produced by and purified from *E. coli* as a GST fusion protein. The purified p27 protein (1 μg) was first incubated with phosphorylated Cdk2-CyclinA (10 μg) in a 5 μl solution containing 20 mM Tris-HCl, pH 7.6, 200 mM NaCl, 10 mM MgCl₂, 2% glycerol, and 3 μM ³²P-ATP at room temperature for 30 min. The purified Cul1-Rbx1 complex (3.5 μg), Cks1 (0.3 μg) and the Skp1-Skp2 complex (2.5 μg) were then added to the phosphorylated p27 sample supplemented with insect-cell-expressed and purified E1 (0.5 μg), UbcH5 E2 (0.6 μg), ubiquitin (12.5 μg), and 5 mM ATP. The mixture was incubated at room temperature for additional 30 minutes and analysed by SDS-polyacrylamide gel electrophoresis (PAGE). For binding analysis, the phosphorylated p27 sample was mixed with 10 μg GST-Cul1-linker-mutant-Rbx1 complex, which was expressed and purified from insect cells and remained bound to 50 μl glutathione 4B resin in 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 5 mM DTT (buffer A). 2.5 μg purified Skp1-Skp2 complex and 0.3 μg Cks1 were added where indicated and the mixture was incubated on ice for 30 min. After washing with 500 μl of buffer A twice, the proteins were eluted with 100 μl of buffer A supplemented with 10 mM reduced glutathione and resolved by SDS-PAGE.

Received 10 January; accepted 28 February 2002.

1. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
2. Deshaies, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* **15**, 435–467 (1999).
3. Koepp, D. M., Harper, J. W. & Elledge, S. J. How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell* **97**, 431–434 (1999).
4. Strohmaier, H. *et al.* Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature* **413**, 316–322 (2001).
5. Koepp, D. M. *et al.* Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science* **294**, 173–177 (2001).
6. Moberg, K. H., Bell, D. W., Wahrer, D. C., Haber, D. A. & Hariharan, I. K. Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature* **413**, 311–316 (2001).
7. Polakis, P. Wnt signaling and cancer. *Genes Dev.* **14**, 1837–1851 (2000).
8. Slingerland, J. & Pagano, M. Regulation of the cdk inhibitor p27 and its deregulation in cancer. *J. Cell Physiol.* **183**, 10–17 (2000).

9. Latres, E. *et al.* Role of the F-box protein Skp2 in lymphomagenesis. *Proc. Natl Acad. Sci. USA* **98**, 2515–2520 (2001).
10. Gstaiger, M. *et al.* Skp2 is oncogenic and overexpressed in human cancers. *Proc. Natl Acad. Sci. USA* **98**, 5043–5048 (2001).
11. Latif, F. *et al.* Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**, 1317–1320 (1993).
12. Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
13. Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
14. Joazeiro, C. A. & Weissman, A. M. RING finger proteins: mediators of ubiquitin ligase activity. *Cell* **102**, 549–552 (2000).
15. Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274 (1996).
16. Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
17. Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 221–230 (1997).
18. Skowyra, D. *et al.* Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science* **284**, 662–665 (1999).
19. Kamura, T. *et al.* Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science* **284**, 657–661 (1999).
20. Ohta, T., Michel, J. J., Schottelius, A. J. & Xiong, Y. ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol. Cell* **3**, 535–541 (1999).
21. Seol, J. H. *et al.* Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev.* **13**, 1614–1626 (1999).
22. Kipreos, E. T. & Pagano, M. The F-box protein family. *Genome Biol.* **1**, 3002.1–3002.7 (2000).
23. Ganioth, D. *et al.* The cell-cycle regulatory protein Cks1 is required for SCF (Skp2)-mediated ubiquitylation of p27. *Nature Cell Biol.* **3**, 321–324 (2001).
24. Spruck, C. *et al.* A CDK-independent function of mammalian Cks1. Targeting of SCF (Skp2) to the CDK inhibitor p27 (Kip1). *Mol. Cell* **7**, 639–650 (2001).
25. Kipreos, E. T., Lander, L. E., Wing, J. P., He, W. W. & Hedgecock, E. M. cul-1 is required for cell cycle exit in *C. elegans* and identifies a novel gene family. *Cell* **85**, 829–839 (1996).
26. Ivan, M. & Kaelin, W. G. Jr The von Hippel-Lindau tumor suppressor protein. *Curr. Opin. Genet. Dev.* **11**, 27–34 (2001).
27. Zachariae, W. *et al.* Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins. *Science* **279**, 1216–1219 (1998).
28. Yu, H. *et al.* Identification of a cullin homology region in a subunit of the anaphase-promoting complex. *Science* **279**, 1219–1222 (1998).
29. Schulman, B. A. *et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature* **408**, 381–386 (2000).
30. Rice, L. M. & Brunger, A. T. Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. *Mol. Cell* **4**, 85–95 (1999).
31. Mathias, N. *et al.* Cdc53p acts in concert with Cdc4p and Cdc34p to control the G1-to-S-phase transition and identifies a conserved family of proteins. *Mol. Cell Biol.* **16**, 6634–6643 (1996).
32. Grossberger, R. *et al.* Characterization of the DOC1/APC10 subunit of the yeast and the human anaphase-promoting complex. *J. Biol. Chem.* **274**, 14500–14507 (1999).
33. Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
34. Chen, A. *et al.* The conserved RING-H2 finger of ROC1 is required for ubiquitin ligation. *J. Biol. Chem.* **275**, 15432–15439 (2000).
35. Read, M. A. *et al.* Nedd8 modification of cul-1 activates SCF (beta(TrCP))-dependent ubiquitination of IkappaBalpha. *Mol. Cell Biol.* **20**, 2326–2333 (2000).
36. Morimoto, M., Nishida, T., Honda, R. & Yasuda, H. Modification of cullin-1 by ubiquitin-like protein Nedd8 enhances the activity of SCF (skp2) toward p27 (kip 1). *Biochem. Biophys. Res. Commun.* **270**, 1093–1096 (2000).
37. Kawakami, T. *et al.* NEDD8 recruits E2-ubiquitin to SCF E3 ligase. *Embo J.* **20**, 4003–4012 (2001).
38. Wu, K., Chen, A., Tan, P. & Pan, Z. Q. The Nedd8-conjugated ROC1-CUL1 core ubiquitin ligase utilizes Nedd8 charged surface residues for efficient polyubiquitin chain assembly catalyzed by Cdc34. *J. Biol. Chem.* **277**, 516–527 (2002).
39. Shirane, M. *et al.* Down-regulation of p27 (Kip1) by two mechanisms, ubiquitin-mediated degradation and proteolytic processing. *J. Biol. Chem.* **274**, 13886–13893 (1999).
40. Patton, E. E. *et al.* Cdc53 is a scaffold protein for multiple Cdc34/Skp1/F-box protein complexes that regulate cell division and methionine biosynthesis in yeast. *Genes Dev.* **12**, 692–705 (1998).
41. Scherer, D. C., Brockman, J. A., Chen, Z., Maniatis, T. & Ballard, D. W. Signal-induced degradation of I kappa B alpha requires site-specific ubiquitination. *Proc. Natl Acad. Sci. USA* **92**, 11259–11263 (1995).
42. Stroschein, S. L., Bonni, S., Wrana, J. L. & Luo, K. Smad3 recruits the anaphase-promoting complex for ubiquitination and degradation of SnoN. *Genes Dev.* **15**, 2822–2836 (2001).
43. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
45. Brunger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
46. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. CCP4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
48. Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J. & Pavletich, N. P. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**, 325–331 (1996).
49. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
50. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank H. Erdument-Bromage for N-terminal sequencing and mass spectroscopic analysis; T. Kamura and Z. Pan for reagents; members of the Pavletich laboratory for discussions; C. Murray for administrative assistance; and the staff of the National Synchrotron Light Source X9B beamline and of the Cornell High Energy Synchrotron Source MacCHESS for help with data collection. B.A.S. was supported by a special fellowship from the Leukemia and Lymphoma Society. This work was supported by the NIH, the Howard Hughes Medical Institute, the Dewitt Wallace Foundation, the Samuel and May Rudin Foundation, the Human Frontier Science Program Organization, the Welch Foundation, the Belfer Foundation and the Irma T. Hirschl Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.P.P. (e-mail: nikola@xray2.mskcc.org). Coordinates have been deposited with the Protein Data Bank under accession codes 1LDD, 1LDJ and 1LDK.

The binary Kuiper-belt object 1998 WW31

Christian Veillet*, Joel Wm. Parker†, Ian Griffin‡, Brian Marsden§, Alain Doressoundiram||, Marc Buie¶, David J. Tholen#, Michael Connelley# & Matthew J. Holman§

* Canada France Hawaii Telescope, Kamuela, Hawaii 96743, USA

† Southwest Research Institute, Boulder, Colorado 80302, USA

‡ Space Telescope Science Institute, Baltimore, Maryland 21218, USA

§ Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

|| Observatoire de Paris, Place Jules Janssen, F-92 195 Meudon Cedex, France

¶ Lowell Observatory, 1400 Mars Hill Road, Flagstaff, Arizona 8600, USA

Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

The recent discovery^{1,2} of a binary asteroid during a spacecraft fly-by generated keen interest, because the orbital parameters of binaries can provide measures of the masses, and mutual eclipses could allow us to determine individual sizes and bulk densities. Several binary near-Earth^{3–5}, main-belt^{6–10} and Trojan¹¹ asteroids have subsequently been discovered. The Kuiper belt—the region of space extending from Neptune (at 30 astronomical units) to well over 100 AU and believed to be the source of new short-period comets¹²—has become a fascinating new window onto the formation of our Solar System since the first member object, not counting Pluto, was discovered in 1992 (ref. 13). Here we report that the Kuiper-belt object 1998 WW31 is binary with a highly eccentric orbit (eccentricity $e \approx 0.8$) and a long period (about 570 days), very different from the Pluto/Charon system, which was hitherto the only previously known binary in the Kuiper belt. Assuming a density in the range of 1 to 2 g cm^{−3}, the albedo of the binary components is between 0.05 and 0.08, close to the value of 0.04 generally assumed for Kuiper-belt objects.

1998 WW31 was discovered at the Kitt Peak National Observatory (KPNO) 4-m telescope in November 1998 by R. Millis and his collaborators and reported¹⁴ as a single Kuiper-belt object (KBO). There were only a few positions on a 42-day long arc following the discovery, and no recovery observations were made during its next opposition. Recovery is a very important step in the study of KBOs: based on positions on only a short arc, the orbit of a KBO is not well constrained and the possibility of losing it after a few years without new observations is very high. The uncertain orbital parameters of lost KBOs make them nearly useless for subsequent analysis of the Kuiper belt. Therefore, 1998 WW31 was included in the list of objects to be observed during a KBO photometry and recovery programme by C.V. and A.D. scheduled in late December 2000 on the Canada-France-Hawaii 3.6-m telescope (CFHT). A first image of the field of 1998 WW31 was taken on 21 December 2000, and two additional images were obtained the following night. A first look at the images immediately after the observations did not reveal the presence of 1998 WW31. But in April 2001 C.V. re-processed all the images of the run and the object was located. Figure 1 shows a composite of the recovery images.

A first analysis of these three images (Fig. 1a) did not show any significant relative motion of the two components. It was unclear from these few images if 1998 WW31 was a true binary object, or if it only appeared to be so, by chance superposition of two non-related KBOs in different orbits but seen close to each other with apparent motions similar enough to be indistinguishable over the course of two nights. In April 2001, the solar elongation angle of 1998 WW31 was too small to allow any observation. Fortunately, we had at our disposal data taken during the previous year, allowing us to look at the object farther in the past to determine whether it was binary. In searching the CFHT archives we found seven images of 1998 WW31 taken on 6 and 7 January 2000 by J. J. Kavelaars and A. Morbidelli.

We generated a new ephemeris for 1998 WW31 including our new observations, and computed the position of the KBO for the dates on which these images were taken. Thus, we were able to find 1998 WW31 on those images, embedded in the halo of a very bright star on the first night and close to another star on the second night, explaining why the other team was not successful in locating the object. In many of these images, 1998 WW31 was clearly elongated or even resolved into two components. The orientation and the separation of the components were different from those seen in our observations nearly a year later (Fig. 1), confirming that 1998 WW31 is a binary KBO.

A further search through the CFHT observing log showed that another set of observations had been taken shortly after our December run at CFHT by H. Aussel for D. Tholen. In the *Minor Planet Electronic Circulars*¹⁵, where our recovery had been published, two observations by M. Buie at KPNO were also reported from November 2000. Subsequently, J. Parker found 1998 WW31 on images he had taken on 29 and 30 November 2000 at KPNO, and examination of these images further confirmed the presence of both components.

We contacted all the observers who had images of 1998 WW31. The ground-based observations since the 1998 WW31 discovery that had adequate seeing to resolve the components covered slightly more than two years in six main epochs. The data are of variable quality. Some of them were acquired under conditions of poor atmospheric seeing. Owing to the spacing of the observations, it was difficult to assess confidently the orbital period, though a long one (around 500 days) was more likely than a short one (150 days). To resolve this uncertainty, we obtained Director's Discretionary Time with the Hubble Space Telescope (HST) to observe 1998 WW31

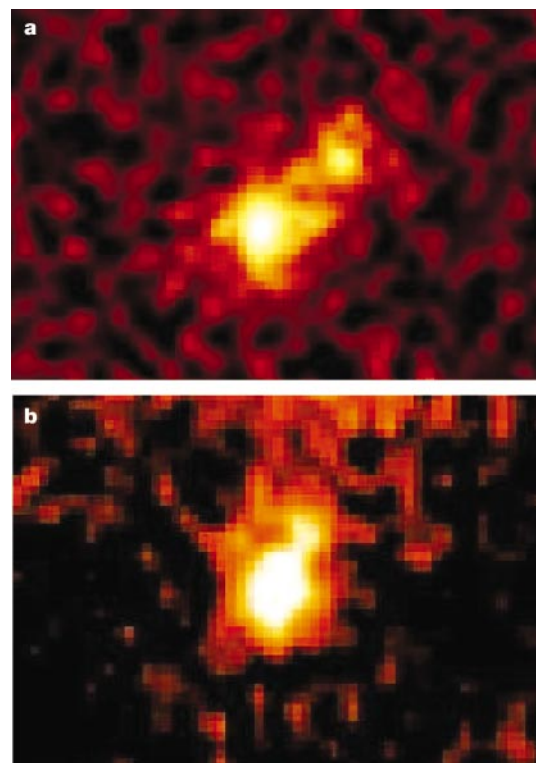


Figure 1 1998 WW31 on **a**, 22 December 2000 (binarity discovery image), and **b**, nearly a year earlier on 7 January 2000. The two images are composites of 3 (**a**) and 4 (**b**) raw images, respectively, taken with CFH12K, the wide-field imaging camera mounted at the prime focus of the Canada-France-Hawaii 3.6-m telescope. 1998 WW31 is clearly a double KBO, although the details of the orbit cannot be determined from these two images alone. The angular separation between the two sources is 1.2 arcsec in December 2000 and 0.8 arcsec in January 2000.

with the Wide-Field Planetary Camera 2 (WFPC2) instrument¹⁶ once a month for three months. The first observations were scheduled at the earliest possible date, early July 2001, when the solar elongation angle of 1998 WW31 was larger than the HST solar avoidance angle. For each of our observations 1998 WW31 was centred on the planetary camera, giving a resolution of 0.046 arcsec pixel⁻¹.

With a separation on the sky of about 0.7 arcsec the two components were easily resolved (Fig. 2). Observations were obtained at three separate epochs: 12 July, 9 August, and 10 September 2001. At each epoch, two exposures were made through each of the F555W, F675W and F814W filters (comparable to the Johnson system V, R and I filters). Data were reduced using the standard WFPC2 pipeline¹⁶. We performed astrometric and photometric measurements with the Image Reduction and Analysis Facility¹⁷ (IRAF) software. In parallel with the HST observations, we obtained ground-based observations from CFHT using the queued service observing mode operating the wide-field imager. With these coordinated observations, we were able to get a well-sampled series of images over three months (usually with excellent seeing conditions at CFHT). From all these positions, the orbit of the secondary component was determined using various weight combinations in the very inhomogeneous data set. The eccentricity was found to be at least 0.5 and not well constrained on the upper

side owing to the lack of positions close to the pericentre. Even an orbit with an eccentricity as high as 0.9 could fit the observations reasonably well.

A moderate eccentricity of around 0.6 would place the date of the pericentre in early 2002, so additional HST visits were obtained on Director's Discretionary Time on 30 December 2001 and 19 January 2002 to refine the ellipticity of the orbit. An additional set of HST observations made on 13 December 2001, obtained as part of a separate programme, were made available to us by K. Noll. These observations, made through filters identical to those used in our observations did however place the image of 1998 WW31 on the wide-field camera, and hence were of resolution half that of our other HST observations.

Figure 3 shows our best fits to the observations. The orbit computation algorithms developed for WW31 were first checked and confirmed with the pair Pluto/Charon, using the HST observations¹⁸. The same algorithms were then applied to the 1998 WW31 pair. The uncertainties assigned to the observations and used for determining the formal uncertainty on the orbital elements, are outlined on Fig. 3. Two sets of orbital elements were computed: the first one uses only the HST observations, and the second one uses the whole set of observations. Results are given in Table 1. The HST observations, which cover one-third of the orbital period, comprise two short arcs (60 and 40 days long, respectively) separated by 100 days, with the pericentre occurring between them. The eccentricity is therefore well determined, but other elements, especially the period and the overall orbital geometry, are dramatically improved when all data are used, in spite of the lower quality of the early observations.

The 1998 WW31 pair is very different from the Pluto/Charon system, as seen on Table 1. The most unusual feature of the orbit is its eccentricity, the largest one found in a binary asteroid. Additional HST observations will allow us to follow the faint component as it goes through the apocentre and to determine better the period and

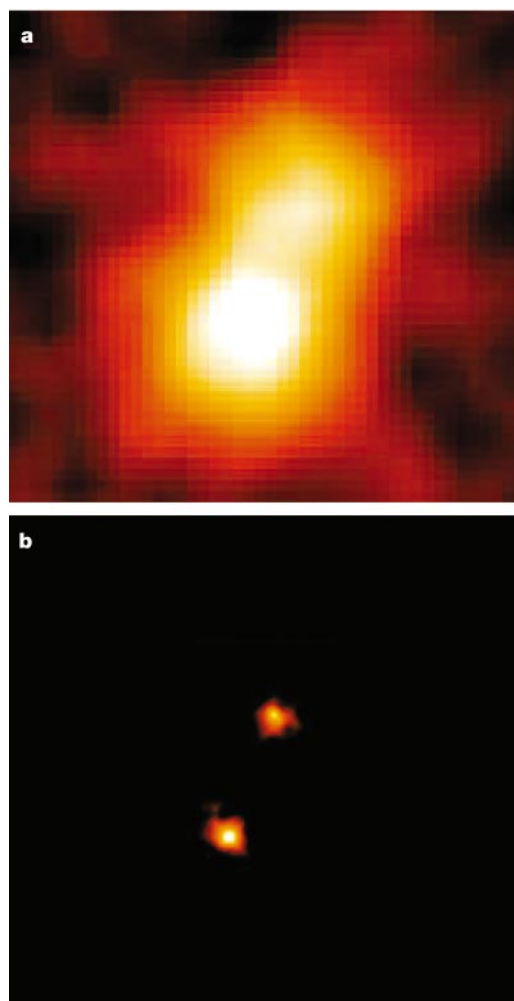


Figure 2 1998 WW31 seen **a**, from the ground at CFHT on 2001 September 12.5 with an excellent seeing of 0.5 arcsec, and **b**, from space with HST on September 9.7. The images are at the same scale. Separation of the components is 0.59 arcsec.

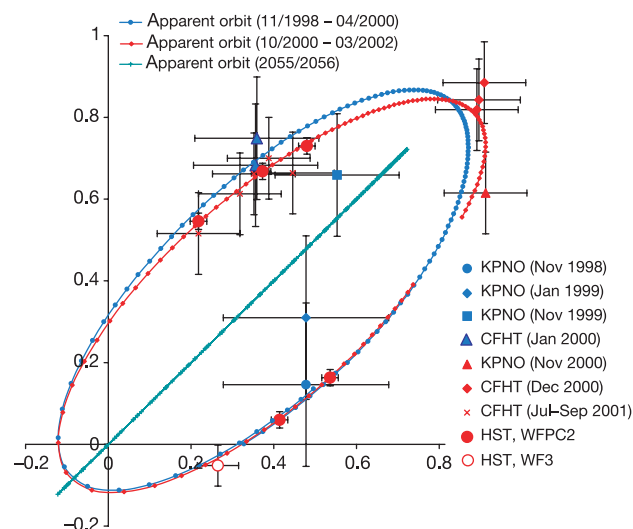


Figure 3 1998 WW31 observations, from the ground and from the Hubble Space Telescope. Coordinates are seconds of arc. North is up and East is to the right. The apparent orbits at the time of the observations are shown with one dot every five days. The apparent orbit in 2055/2056 (green line) is also plotted to show that mutual eclipses could happen at some point nearly 50 years from now (with a few years uncertainty). The following uncertainties were assigned on the relative apparent rectangular coordinates of the faint component relative to the main one (in arcsec): 0.02 for the HST WFPC2 observations (red circles), 0.05 for the WF3 observation (open red circle; pair poorly resolved), 0.1 for the data close to the apocentre (red triangle and diamonds), 0.2 for the two earliest ground-based observations (blue circles and diamonds), and 0.15 for all the others.

Table 1 Relative orbital elements of 1998 WW31 binary system

	All observations	HST only	Charon
Period (days)	574 (10)	521 (133)	6.3872
Semi-major axis (km)	22,300 (800)	21,200 (2,400)	19,366
Eccentricity	0.817 (0.05)	0.800 (0.10)	0.0076
Inclination (degrees)	41.7 (7)	43.8 (10)	96.16
Node longitude (degrees)	94.3 (8)	94.6 (10)	223.0
Pericentre longitude (degrees)	253.8 (7)	251.8 (9)	
Mean longitude at epoch (degrees)	43.4 (4)	48.7 (16)	
Epoch (Julian date)	2,452,300.5	2,452,300.5	
R band magnitudes			
Pair	23.6		
A	24.2		
B	24.6		
Same albedo			
Diameter ratio	1.2		
Mass ratio	1.74		
Same density 2.0 g cm^{-3} (similar to Pluto)			
Diameters (A, B)	118 km, 98 km		
Albedo	0.086 (0.007)		
Same density 1.5 g cm^{-3}			
Diameters (A, B)	129 km, 108 km		
Albedo	0.071 (0.006)		
Same density 1.0 g cm^{-3}			
Diameters (A, B)	148 km, 123 km		
Albedo	0.054 (0.005)		

The orbital elements (mean equator and equinox of J2000.0) of the secondary component, B, of the 1998 WW31 system with respect to its primary, A. (Formal uncertainties in brackets are based on the uncertainties assigned to the individual measurements used as outlined in Fig. 3.) Also shown are the physical properties of the 1998 WW31 components derived from the mass of the system as determined from the orbital parameters (period and semi-major axis), for various assumptions on their density. The orbital elements based only on the HST data are shown to stress the importance of the ground observations in spite of their poorer quality. Charon's elements are from ref. 14.

semi-major axis; however, even the values presented in this paper are accurate enough to lead to a good estimate of the mass of the system.

The orbit is not determined well enough to predict with precision when mutual eclipses could happen. With the present solution, eclipses could take place in 2055 or 2056. The HST observations scheduled for the next opposition will allow a better prediction of when mutual events will occur.

Using both ground-based and HST observations, the two components are found to have a magnitude difference in R band of 0.4 and a total magnitude of 23.6. If we assume that they have the same albedo, their diameter ratio is 1.2. If they have the same density, it is possible to get the mass of each component. The diameters can be estimated assuming a given density. An albedo can then be derived from the observed magnitude and the estimated diameter. Table 1 summarizes the results obtained assuming various densities from 1 to 2 g cm^{-3} (Pluto's density is about 2 g cm^{-3}). The associated albedo values are in the range 0.05 to 0.08. The albedo of cometary nuclei, 0.04 (ref. 19), is the value generally assumed and used for KBO size estimation; however, if we use that low albedo value for 1998 WW31, then the KBO would have a very low density of roughly 1 g cm^{-3} , considerably less than the density of Pluto.

The announcement of the binarity of 1998 WW31 was the first of what has become a series of binary KBO discoveries. Within less than a year after our announcement of the binarity of 1998 WW31, six other KBOs have been discovered to be binaries: 2001 QT297 (ref. 20) and 2001 QW322 (ref. 21) using ground-based observations, and 1999 TC36 (ref. 22), (26308) 1998 SM165 (ref. 23), 1997 CQ29 (ref. 24), and 2000 CF105 (ref. 25) using HST. We now know of seven binary systems in a sample of nearly 600 objects. Although we have to be careful when dealing with small numbers, binarity is definitely not uncommon in the Kuiper belt, comprising at least 1% of the currently known KBO population. □

Received 5 December 2001; accepted 27 February 2002.

1. Chapman, C. R. *et al.* Discovery and physical properties of Dactyl, a satellite of asteroid 243 Ida. *Nature* **374**, 783–785 (1995).
2. Belton, M. J. S. *et al.* The discovery and orbit of 1993 (243)1 Dactyl. *Icarus* **120**, 185–199 (1995).
3. Margot, J. L. *et al.* Satellites of minor planets. *IAU Circ.* 7503 (2000).
4. Benner, L. A. M. *et al.* 1999_KW4. *IAU Circ.* 7632 (2001).

5. Nolan, M. C. *et al.* 2000_UG11. *IAU Circ.* 7632 (2000).
6. Merline, W. J. *et al.* Discovery of a moon orbiting the asteroid 45 Eugenia. *Nature* **401**, 565–567 (1999).
7. Merline, W. J. *et al.* Discovery of companions to asteroids 762 Pulcova and 90 Antiope by direct imaging. DPS Meeting 32, abstr. no. 13.06 (American Astronomical Society, 2000).
8. Brown, M. E. & Margot, J. L. S/2001 (87) 1. *IAU Circ.* 7588 (2001).
9. Merline, W. J. *et al.*; Margot, J. L. & Brown, M. E. S/2001 (22) 1. *IAU Circ.* 7703 (2001).
10. Merline, W. J. *et al.* S/2002 (3749) 1. *IAU Circ.* 7827 (2002).
11. Merline, W. J. *et al.* S/2001 (617) 1. *IAU Circ.* 7741 (2001).
12. Duncan, M., Quinn, T. & Tremaine, S. The origin of short-period comets. *Astrophys. J.* **328**, L69–L73 (1988).
13. Jewitt, D. C. & Luu, J. X. 1992 QB1. *IAU Circ.* 5611 (1992).
14. Millis, R. L. *et al.* 1998 WW31. *Minor Planet Electron. Circ.* 1999-B24 (2001).
15. Veillet, C. *et al.* 1998 WW31. *Minor Planet Electron. Circ.* 2001-G29 (2001).
16. Holtzman, J. *et al.* The performance and calibration of WFPC2 on the Hubble Space Telescope. *Publ. Astron. Soc. Pacif.* **107**, 156–178 (1995).
17. Image Reduction and Analysis Facility. (<http://iraf.noao.edu>).
18. Tholen, D. J. & Buie, M. W. The orbit of Charon. *Icarus* **125**, 245–260 (1997).
19. Jewitt, D., Ausel, H. & Evans, A. The size and albedo of the Kuiper-belt object (20000) Varuna. *Nature* **411**, 446–447 (2001).
20. Elliot, J. L., Kern, S. D., Osip, D. J. & Burles, S. M. 2001 QT₂₉₇. *IAU Circ.* 7733 (2001).
21. Kavelaars, J. J. *et al.* 2001 WQ₃₂₂. *IAU Circ.* 7749 (2001).
22. Trujillo, C. A. & Brown, M. E. 1999 TC₃₆. *IAU Circ.* 7787 (2001).
23. Brown, M. E. & Trujillo, C. A. (26308) 1998 SM₁₆₅. *IAU Circ.* 7807 (2001).
24. Noll, K. *et al.* 1997 CQ₂₉. *IAU Circ.* 7824 (2002).
25. Noll, K. *et al.* 2000 CF₁₀₅. *IAU Circ.* 7857 (2002).

Acknowledgements

C.V., A.D., D.J.T. and M.C. are visiting astronomers at the Canada-France-Hawaii Telescope (CFHT), operated by the National Research Council of Canada, the Centre National de la Recherche Scientifique of France and the University of Hawaii. J.Wm.P. and M.B. are visiting astronomers at the Kitt Peak National Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation. This work, supported by funding from NASA, is partly based on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA under a NASA contract.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.V. (e-mail: veillet@cfht.hawaii.edu).

Electrical detection of spin precession in a metallic mesoscopic spin valve

F. J. Jedema, H. B. Heersche, A. T. Filip, J. J. A. Baselmans & B. J. van Wees

Department of Applied Physics and Materials Science Center, University of Groningen, Nijenborgh 4.13, 9747 AG Groningen, The Netherlands

To study and control the behaviour of the spins of electrons that are moving through a metal or semiconductor is an outstanding challenge in the field of 'spintronics', where possibilities for new electronic applications based on the spin degree of freedom are currently being explored^{1–5}. Recently, electrical control of spin coherence⁶ and coherent spin precession during transport⁷ was studied by optical techniques in semiconductors. Here we report controlled spin precession of electrically injected and detected electrons in a diffusive metallic conductor, using tunnel barriers in combination with metallic ferromagnetic electrodes as spin injector and detector. The output voltage of our device is sensitive to the spin degree of freedom only, and its sign can be switched from positive to negative, depending on the relative magnetization of the ferromagnetic electrodes. We show that the spin

direction can be controlled by inducing a coherent spin precession caused by an applied perpendicular magnetic field. By inducing an average precession angle of 180° , we are able to reverse the sign of the output voltage.

In our experiment we use a mesoscopic spin valve (Fig. 1a), where a cobalt ferromagnetic electrode (Co1) injects spin-polarized electrons into an aluminium (Al) strip via a tunnel barrier. At a distance L from the injector a second cobalt electrode (Co2) is placed, which detects spin-polarized electrons in the Al strip through a tunnel barrier. The presence of the tunnel barriers is crucial, as they provide a high spin dependent resistance, which enhances the spin polarization of the injected current flowing into the Al strip^{8–10}. In addition, the barriers cause the electrons, once injected, to have a negligible probability of losing their spin information by escaping into the Co electrodes. During the time of travel from injector to detector, the spin direction of the electrons can therefore only be altered by (random) spin flip scattering processes in the Al strip or, in the presence of an external magnetic field, by coherent precession. Here we experimentally demonstrate both processes by measuring the amplitude of the spin signal, first as a function of the Co electrode spacing L and second as a function of an applied perpendicular magnetic field. A related method of probing spin injection and detection has been reported by Johnson and Silsbee³. However, the reduction of the sample dimensions by 3 orders of magnitude and the introduction of tunnel barriers enables us to observe a clear sign reversal of the output voltage V due to coherent precession, and allows us to make direct comparison with theory.

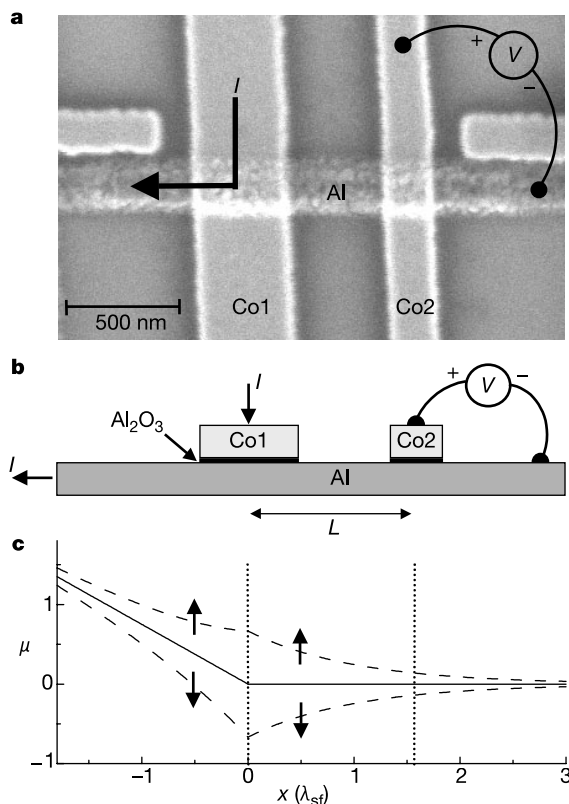


Figure 1 Geometry of our spin valve device. **a**, Scanning electron microscope image of a device with a cobalt (Co) electrode spacing of $L = 650$ nm. Current is sent from Co1 into the Al strip. The voltage is measured between Co2 and the right side of the Al strip. **b**, Device cross-section. **c**, The spatial dependence of the spin-up and spin-down electrochemical potentials (μ , dashed lines) in the Al strip. The solid lines indicate the electrochemical potential (voltage) of the electrons in the absence of spin injection. λ_{sf} spin flip length.

We made a batch of 10 devices with L ranging from 550 to 1,350 nm, using a suspended shadow mask evaporation process¹¹ and electron beam lithography for patterning (Fig. 1a). In the first step, an Al strip with a thickness of 50 nm and a width of 250 nm is evaporated on a thermally oxidized silicon substrate by electron-gun evaporation. Next, the Al strip is exposed to an oxygen (O_2) environment of 5×10^{-3} mbar for 10 minutes, producing a thin aluminium oxide (Al_2O_3) layer. In the third step, without breaking the vacuum, we evaporate two ferromagnetic Co electrodes with sizes of $0.4 \times 4 \mu m^2$ (Co1) and $0.2 \times 12 \mu m^2$ (Co2) and a thickness of 50 nm. Two Al/ Al_2O_3 /Co tunnel junctions are thus formed at the overlap of the Co electrodes and the Al strip (Fig. 1b). The conductivity of the Al film was measured to be $\sigma_{Al} = 1.1 \times 10^7 \Omega^{-1} m^{-1}$ at room temperature and $\sigma_{Al} = 1.7 \times 10^7 \Omega^{-1} m^{-1}$ at 4.2 K. The resistance of the tunnel barriers was determined to be typically 600 Ω of the Co1 electrode and 1,200 Ω for the Co2 electrode at room temperature, both increasing by 10% at 4.2 K. Different geometric aspect ratios of Co1 and Co2 are used to obtain different coercive fields. This allows us to control their relative magnetization configuration (parallel/antiparallel) by sweeping an applied magnetic field B , directed parallel to their long axes⁴.

The spin polarization P of the current I injected from the Co1 electrode into the Al strip is determined by the ratio of the different spin-up and spin-down tunnel barrier resistances $R_\uparrow$ and $R_\downarrow$, and in first order can be written¹² as $P = (N_\uparrow - N_\downarrow)/(N_\uparrow + N_\downarrow)$. Here $N_\uparrow(N_\downarrow)$ is the spin-up (spin-down) density of states at the Fermi level of the electrons in the Co electrodes. The injected spin current

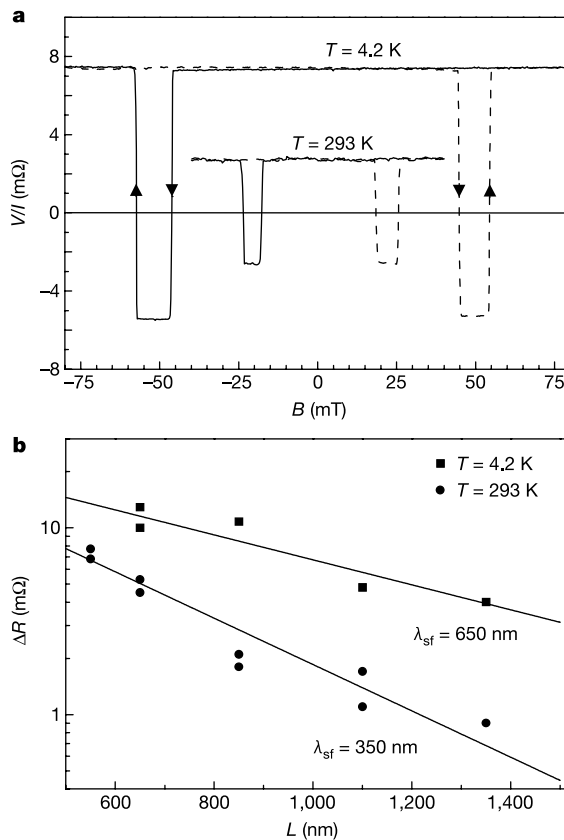


Figure 2 Spin valve effect. **a**, Output signal V/I as a function of the in-plane magnetic field B for a sample with a Co electrode spacing $L = 650$ nm at $T = 4.2$ K and room temperature. The solid (dashed) lines correspond to the negative (positive) sweep direction. **b**, The dependence of the spin-dependent resistance ΔR on the Co electrode spacing L at $T = 4.2$ K and room temperature. The solid squares represent data taken at $T = 4.2$ K, the solid circles are taken at room temperature. The solid lines represent the best fits based on equation (1).

causes the densities (or electrochemical potentials) of the spin-up and spin-down electrons in the Al strip to become unequal (Fig. 1c). This unbalance is transported to the Co2 detector electrode by diffusion, and can therefore be detected. Owing to the spin-dependent tunnel barrier resistances, the Co2 electrode detects a weighted average of the two spin densities, which causes the detected output voltage V to be proportional to P^2 .

Figure 2a shows a typical output signal V/I as a function of an in-plane magnetic field B , directed parallel to the long axes of Co1 and Co2, taken at room temperature and 4.2 K. The measurements are performed by standard a.c. lock-in techniques, using a current $I = 100 \mu\text{A}$. Sweeping the magnetic field from negative to positive, a sign reversal of the output signal is observed, when the magnetiza-

tion of Co1 flips at 19 mT (room temperature) and 45 mT (4.2 K), and the device switches from a parallel to antiparallel configuration. When the magnetization of Co2 flips at 25 mT (room temperature) and 55 mT (4.2 K), the magnetizations are parallel again, but now point in the opposite direction. The fact that the output signal switches symmetrically around zero indicates that this experiment is sensitive to the spin degree of freedom only.

We have calculated the expected magnitude of the output signal V/I as a function of the Co electrode spacing L by solving the spin coupled diffusion equations for the spin-up and spin-down electrons in the Al strip^{13–15}. Taking into account the fact that the tunnel barrier resistances are much larger than the resistance of the Al strip over a spin flip length, we obtain:

$$\frac{V}{I} = \pm \frac{1}{2} P^2 \frac{\lambda_{\text{sf}}}{\sigma_{\text{Al}} A} \exp(-L/\lambda_{\text{sf}}) \quad (1)$$

where $\lambda_{\text{sf}} = \sqrt{D\tau_{\text{sf}}}$ is the spin flip length, A the cross-sectional area, D the diffusion constant, and τ_{sf} the spin flip time of the Al strip. The positive (negative) sign corresponds to a parallel (antiparallel) magnetization configuration of the Co electrodes.

Figure 2b shows the measured spin dependent resistance $\Delta R = \Delta V/I$ as a function of L , where ΔV is the output voltage difference between parallel and antiparallel configuration. By fitting the data to equation (1), we find $P = 0.11 \pm 0.02$ at both 4.2 K and room temperature, $\lambda_{\text{sf}} = 650 \pm 100 \text{ nm}$ at 4.2 K and $\lambda_{\text{sf}} = 350 \pm 50 \text{ nm}$ at room temperature. The diffusion constant D is calculated using the Einstein relation $\sigma_{\text{Al}} = e^2 N_{\text{Al}} D$, where e is the electron charge and $N_{\text{Al}} = 2.4 \times 10^{22}$ states per eV per cm^3 is the density of states of Al at the Fermi energy¹⁶. Using $D = 4.3 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ at 4.2 K and $D = 2.7 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ at room temperature, we obtain $\tau_{\text{sf}} = 100 \text{ ps}$ at 4.2 K and $\tau_{\text{sf}} = 45 \text{ ps}$ at room temperature. These values are in good agreement with those reported in the literature^{3,17–20}.

Having determined the parameters P , λ_{sf} and D , we are now ready to study spin precession of the electron spin during its diffusion time t between Co1 and Co2. In an applied field $B_{\perp}$, perpendicular to the substrate plane, the injected electron spins in the Al strip precess around an axis parallel to $B_{\perp}$. This alters the spin direction by an angle $\varphi = \omega_L t$, where $\omega_L = g\mu_B B_{\perp}/\hbar$ is the Larmor frequency, g is the g -factor of the electron (~ 2 for Al), μ_B is the Bohr magneton and $\hbar$ is Planck's constant divided by 2π . Because the Co2 electrode detects the projection of the spin direction φ onto its own magnetization direction (0 or π), the contribution of an electron to

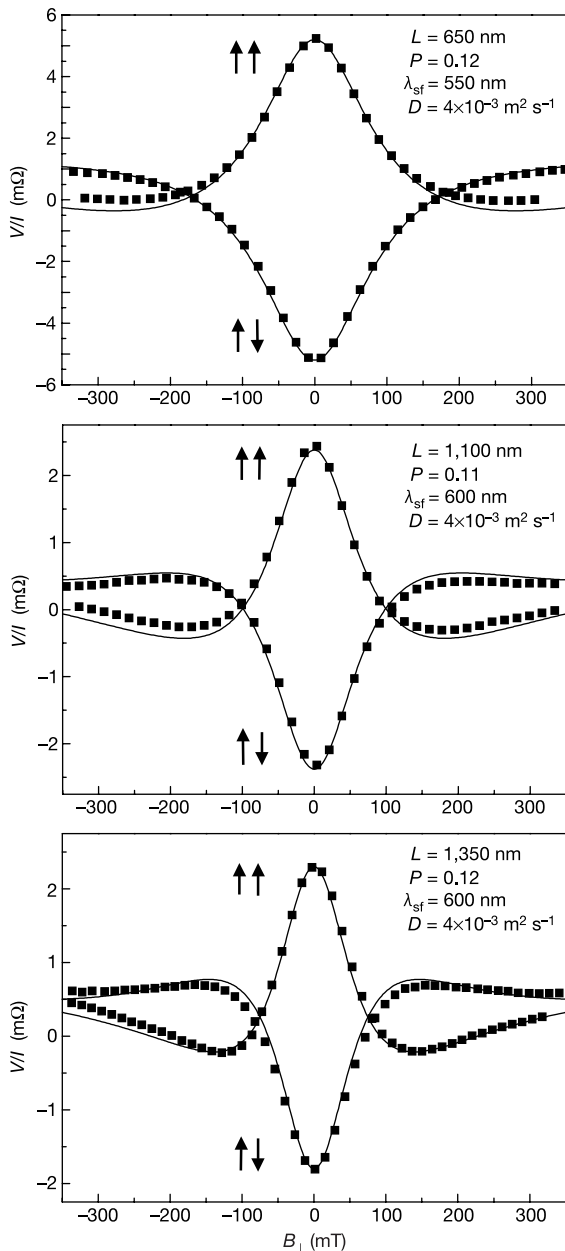


Figure 3 Modulation of the output signal V/I due to spin precession as a function of a perpendicular magnetic field $B_{\perp}$, for $L = 650 \text{ nm}$, $L = 1,100 \text{ nm}$ and $L = 1,350 \text{ nm}$. The solid squares represent data taken at $T = 4.2 \text{ K}$, whereas the solid lines represent the best fits based on equations (2) and (3). The arrows indicate the relative magnetization configuration (parallel/antiparallel) of the Co electrodes. P , spin polarization; D , diffusion constant.

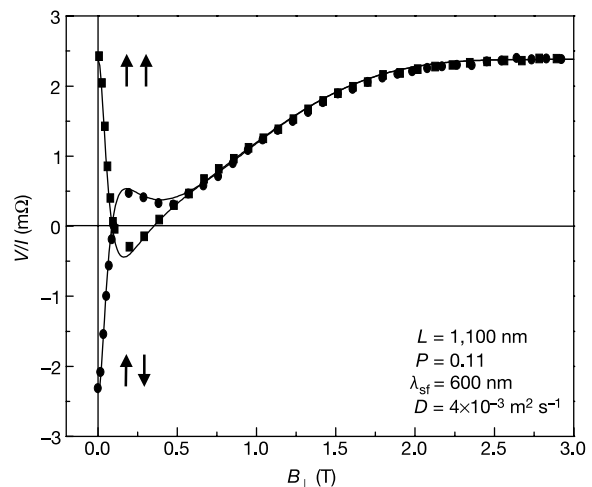


Figure 4 Modulation of the output signal V/I as a function of a perpendicular magnetic field $B_{\perp}$ up to 3 T, for $L = 1,100 \text{ nm}$. The solid squares/circles represent data taken at $T = 4.2 \text{ K}$, whereas the solid lines represent the best fit based on equations (2) and (3). The arrows indicate the relative magnetization configuration (parallel/antiparallel) of the Co electrodes.

the output voltage V is proportional to $\pm \cos(\varphi)$. However, in an (infinite) diffusive conductor the diffusion time t from Co1 to Co2 has a broad distribution $P(t) = [1/\sqrt{4\pi Dt}] \exp[-L^2/(4Dt)]$, where $P(t)$ is proportional to the number of electrons that, once injected at the Co1 electrode ($x = 0$), arrive at the Co2 electrode ($x = L$) after a diffusion time t . The output voltage V at the Co2 detector electrode as a function of $B_{\perp}$ is calculated by summing all contributions of the electron spins over all diffusion times t . We obtain:

$$V(B_{\perp}) = \pm I \frac{P^2}{e^2 N_{\text{Al}} A} \int_0^{\infty} P(t) \cos(\omega_L t) \exp(-t/\tau_{\text{sf}}) dt \quad (2)$$

The exponential factor in equation (2) describes the effect of the spin flip scattering. For $\omega_L = 0$, equation (2) reduces to equation (1). We note that equation (2) can be evaluated analytically, and we have verified that it yields the same result as obtained by Johnson and Silsbee, who explicitly solved the Bloch equations with the appropriate boundary conditions^{21,22}.

At large $B_{\perp}$, the magnetization direction of the Co electrodes is tilted out of the substrate plane with an angle ϑ . When we include this effect we calculate:

$$V(B_{\perp}, \vartheta) = V(B_{\perp}) \cos^2(\vartheta) + |V(B_{\perp} = 0)| \sin^2(\vartheta) \quad (3)$$

Equation (3) shows that with increasing ϑ (from zero), the precession signal is reduced and a positive background output signal appears. For $\vartheta = 0$ equation (3) reduces to equation (2). In the limit that $\vartheta = \pi/2$, the magnetization of the Co electrodes is perpendicular to the substrate plane and parallel to $B_{\perp}$. No precession now occurs, and the full output signal $|V(B_{\perp} = 0)|$ is recovered. The angle ϑ has been determined independently as a function of $B_{\perp}$ by measuring the anisotropic magnetoresistance of the Co electrodes²³.

In Fig. 3 we plot the measured output signal V/I at 4.2 K, as a function of $B_{\perp}$ for $L = 650$ nm, $L = 1,100$ nm and $L = 1,350$ nm. Before the measurement an in-plane magnetic field B directed parallel to the long axes of Co electrodes is used to prepare the magnetization configuration of the Co electrodes. For a parallel (antiparallel) configuration we observe an initial positive (negative) signal, which drops in amplitude as $B_{\perp}$ is increased from zero field. This is called the Hanle effect in ref. 3. The two curves cross where the average angle of precession is about 90° and the output signal is close to zero. As $B_{\perp}$ is increased beyond this field, we observe that the output signal changes sign and reaches a minimum (maximum) when the average angle of precession is about 180° , thereby effectively converting the injected spin-up population into a spin-down and vice versa. We have fitted the data with equations (2) and (3), as shown in Fig. 3. We find for all measured samples that the best-fit parameters P , λ_{sf} and D are very close to those independently obtained from the length dependence measurements (Fig. 2).

As already visible in Fig. 3, for $B > 200$ mT an asymmetry is observed between the parallel and antiparallel curves. This is due to the fact that magnetization of the Co electrodes does not remain in the substrate plane. In Fig. 4 we plot the measured output signal V/I at $T = 4.2$ K for $L = 1,100$ nm up to $B_{\perp} = 3$ T, together with the calculated curve, using P , λ_{sf} and D as obtained from the best fit in Fig. 3. The data are in close agreement with equation (3), and show a suppression of the precessional motion of the electron spin. The full magnitude of the output signal is recovered at large $B_{\perp}$, when $\vartheta = \pi/2$ and no precession takes place. Preliminary results show that precession effects similar to those shown in Fig. 3 can also be obtained at room temperature.

We believe that the system we report here, with its unique sensitivity to the spin degree of freedom, will make possible detailed studies of a variety of spin-dependent transport phenomena. $\square$

Received 18 January; accepted 21 February 2002.

1. Prinz, G. A. Magnetoelectronics. *Science* **282**, 1660–1663 (1998).

- Wolf, S. A. *et al.* Spintronics: A spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
- Johnson, M. & Silsbee, R. H. Interfacial charge-spin coupling: Injection and detection of spin magnetization in metals. *Phys. Rev. Lett.* **55**, 1790–1793 (1985).
- Jedema, F. J., Filip, A. T. & van Wees, B. J. Electrical spin injection and accumulation at room temperature in an all-metal mesoscopic spin valve. *Nature* **410**, 345–348 (2001).
- Hernando, D. H., Nazarov, Yu. V., Brataas, V. & Bauer, G. E. W. Conductance modulation by spin precession in noncollinear ferromagnet normal-metal ferromagnet systems. *Phys. Rev. B* **62**, 5700–5712 (2000).
- Salis, G. *et al.* Electrical control of spin coherence in semiconductor nanostructures. *Nature* **414**, 619–622 (2001).
- Kikkawa, J. M. & Awschalom, D. D. Lateral drag of spin coherence in gallium arsenide. *Nature* **397**, 139–141 (1999).
- Schmidt, G. *et al.* Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor. *Phys. Rev. B* **62**, R4790–R4793 (2000).
- Rashba, E. I. Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem. *Phys. Rev. B* **62**, R16267–R16270 (2000).
- Fert, A. & Jaffrès, H. Conditions for efficient spin injection from a ferromagnetic metal into a semiconductor. *Phys. Rev. B* **64**, 184420–184428 (2001).
- Jackel, L. D., Howard, R. E., Hu, E. L., Tennant, D. M. & Grabbe, P. 50-nm silicon structures fabricated with trilevel electron beam resist and reactive-ion etching. *Appl. Phys. Lett.* **39**, 268–270 (1981).
- Juliere, M. Tunneling between ferromagnetic films. *Phys. Lett. A* **54**, 224–227 (1975).
- Van Son, P., van Kempen, H. & Wyder, P. Boundary resistance of the ferromagnetic-nonferromagnetic metal interface. *Phys. Rev. Lett.* **58**, 2271–2273 (1987).
- Jedema, F. J., Nijboer, M. S., Filip, A. T. & van Wees, B. J. Spin injection and spin accumulation in permalloy-copper mesoscopic spin valves. Preprint cond-mat/0111092 at (<http://xxx.lanl.gov>) (2002); *Phys. Rev. B* (submitted).
- Johnson, M. Spin accumulation in gold films. *Phys. Rev. Lett.* **70**, 2142–2145 (1993).
- Papaconstantopoulos, D. A. *Handbook of the Band Structure of Elemental Solids* (Plenum, New York, 1986).
- Meservey, R. & Tedrow, P. M. Surface relaxation times of conduction-electron spins in superconductors and normal metals. *Phys. Rev. Lett.* **41**, 805–808 (1978).
- Monod, P. & Béné, F. Conduction-electron spin flip by phonons in metals: Analysis of experimental data. *Phys. Rev. B* **19**, 911–916 (1979).
- Grimaldi, C. & Fulde, P. Spin-orbit scattering effects on the phonon emission and absorption in superconducting tunneling junctions. *Phys. Rev. Lett.* **77**, 2550–2553 (1996).
- Fabian, J. & Das Sarma, S. Phonon-induced spin relaxation of conduction electrons in aluminium. *Phys. Rev. Lett.* **83**, 1211–1214 (1999).
- Johnson, M. & Silsbee, R. H. Coupling of electronic charge and spin at a ferromagnetic-paramagnetic metal interface. *Phys. Rev. B* **37**, 5312–5325 (1988).
- Johnson, M. & Silsbee, R. H. Spin-injection experiment. *Phys. Rev. B* **73**, 5326–5335 (1988).
- Rijks, Th. G. S. M., Coehoorn, R., de Jong, M. J. M. & de Jonge, W. J. M. Semiclassical calculations of the anisotropic magnetoresistance of NiFe-based thin films, wires, and multilayers. *Phys. Rev. B* **51**, 283–291 (1995).

Acknowledgements

We thank the Stichting Fundamenteel Onderzoek der Materie (FOM) and NEDO (project 'nano-scale control of magnetoelectronics for device applications') for support.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.J.J. (e-mail: jedema@phys.rug.nl).

Rapid electroplating of insulators

Vincent Fleury*, Wesley A. Watters*, Levy Allam† & Thierry Devers†

* Laboratoire de Physique de la Matière Condensée, Ecole Polytechnique/CNRS, 91128 Palaiseau cedex, France

† Laboratoire de Physique Electronique, IUT de Chartres, 1 Place Mendès-France, 28000 Chartres, France

Electrochemical techniques for depositing metal films and coatings¹ have a long history^{2–5}. Such techniques essentially fall into two categories, with different advantages and disadvantages. The first, and oldest, makes use of spontaneous redox reactions to deposit a metal from solution, and can be used on both insulating and metallic substrates. But the deposition conditions of these processes are difficult to control *in situ*, in part because of the

variety of salts and additives present in the solution. The second approach—electroplating—uses an electric current to reduce metal ions in solution, and offers control over the quantity (and, to some extent, grain size) of deposited metal. But application of this technique has hitherto been restricted to conducting substrates. Here we describe an electroplating technique that permits coating of insulating substrates with metals having controlled grain size, thickness and growth speed. The basis of our approach is the progressive outward growth of the metal from an electrode in contact with the substrate, with the cell geometry chosen so that the electron current providing the reduction passes through the growing deposit. Such an approach would normally form dendritic or powdery deposits, but we identify a range of conditions in which uniform films rapidly form.

Here we describe an electrochemical cell and a set of conditions that enable electrodeposition of a metal coating on insulating substrates⁶. This process allows deposition of a covering film of controlled grain size, thickness and growth speed. This process is based on recent progress in out-of-equilibrium physics, and it makes possible the deposition of a uniform coating in what is regarded as the least useful regime of electrochemical growth.

Deposition of metals by electroplating is known to produce compact metal at low current densities^{7,8}. Then, as the deposition current is raised, the deposit becomes (as a function of current, and in the absence of levelling or grain refining additives) rough, dendritic, and eventually powdery. This is a limiting factor in industry¹. In the context of out-of-equilibrium physics, much attention has been paid to dendrites: the interest is in pattern formation^{9–11}. The specific case of electrochemical growth from binary electrolytes has been extensively studied over the past 15 years^{12–18}. A new theory has been put forward by Chazalviel¹⁶, which correctly predicts the growth speed, deposition rates and concentration fields in the electrolyte, around a growing dendritic deposit. The theory predicts the existence of a large electric field at the tip of the deposit, and a steady-state growth speed equal to the recession speed of the anions. We have verified these predictions in the case of free-floating, almost two-dimensional dendrites^{14,17}, and it was independently found by Melrose *et al.*¹⁴. But one problem with these experiments is that the deposit cannot be taken out of the cell. This is why one of us has proposed a new way of depositing dendrites¹⁸ which adhere to a glass substrate. The growth regime of these dendrites is given by Chazalviel's model too¹⁹, which relates deposition rates and growth speeds.

The growth speed of the deposit is given by $-\mu_a E$, where μ_a is the mobility of the anion, and E the electric field. The thickness t_d of the deposit is predicted to be:

$$t_d = [t_{\text{cell}}(z_a \mu_a + z_c \mu_c) / z_a \mu_a] (C_{\text{solution}} / C_{\text{bulkmetal}})$$

where z_a and z_c are respectively the anionic and cationic charge,

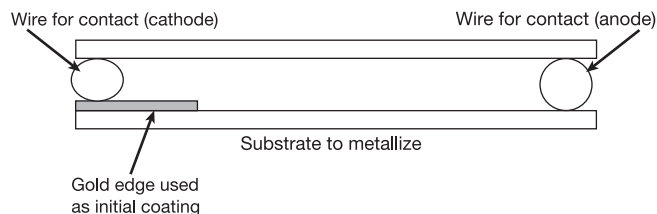


Figure 1 Cell for deposition on a flat surface. A thin layer of electrolyte is squeezed between two glass plates, one of which will be metallized. A gold edge that will serve as the starting point for the metallization is deposited on the glass plate to be metallized. An anode is put at the other end, which also serves as a separator. An additional flash of gold, 20 Å thick and non-conducting, is deposited on the surface (this is not truly necessary, but it helps in rendering the deposition reproducible). The deposition starts from the edge, and invades progressively the entire surface to be metallized.

C_{solution} is the concentration of salt in the solution and $C_{\text{bulkmetal}}$ the concentration of a bulk piece of metal. That it, as all the incoming cations get deposited, they will make a thickness proportional to the ratio $C_{\text{solution}}/C_{\text{bulkmetal}}$ ($C_{\text{bulkmetal}} = 97 \text{ mol l}^{-1}$ for Ag, 141 for Cu, 49 for Sn). In terms of Faraday's law, the total reaction rate for complete metal deposition is I/zF . Because, in the binary electrolyte, metal and counterion transport are dominant in the bulk over other processes:

$$I = (z_a C_{\text{solution}} \mu_a E + z_c C_{\text{solution}} \mu_c E) S$$

where S is the cell cross-section area ($t_{\text{cell}} W$). The cationic deposition rate i_c of a thin film of thickness t_d and front width W advancing at a constant speed $V = \mu_a E$ is $i_c = z_c \mu_a E (t_d W) C_{\text{bulkmetal}}$. Therefore, equating this to the deposition rate in Faraday's law gives the expected thickness if the current efficiency is 100%. These predictions have been verified for surface deposition of copper dendrites¹⁹ in thin cells.

Now we report the result that this set-up makes it possible to deposit a continuous covering film, growing with the same characteristics (thickness, growth speed). There exists a window of conditions, inside the regime of powdery deposition, in which the deposit makes a covering film. This results in a metallization technique, allowing the coating of insulators. Moreover, the working conditions correspond to very rapid deposition, because the coating progresses at speeds up to hundreds of micrometres per second, much quicker than the dendritic regime. As a consequence, we estimate that roughly one linear metre per hour could be deposited.

The set-up has been described in part in ref. 19 (see Fig. 1). The key feature is to start deposition at one end of the surface, at a cathode that is in contact with it. Then, a metal film progressively

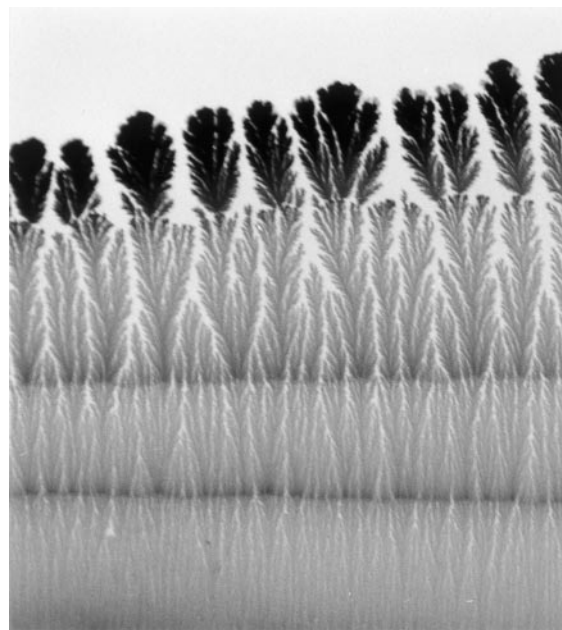


Figure 2 The different overall morphologies for the specific case of copper. For the higher currents (at the start of the deposit), the copper covers the entire substrate. As the current is decreased (in steps), the deposit becomes more and more dendritic, and less space-filling. For very low currents, the voids between dendrites are very large and the deposit is not at all covering. The successive currents are 20 μA , 10 μA , 8 μA , 6 μA : the cell was 15 μm thick only, and the solution is 0.02 M copper sulphate. For the sake of clarity, we did this series of experiments in the same cell (hence the same electrolyte and the same geometry). When doing several growth speeds in the same cell, it is necessary to go from the more uniform to the more dendritic morphology to form a good sample. The other way around implies that the smoother deposit will start at the tips of the very irregular dendrites, which is not favourable.

invades the surface, until it is fully coated. This film wets the surface and adheres to it. This makes it possible to use electroplating to coat an insulator, as the edge of the invading film serves as a cathode for further growth of the film itself. If too low a current density is used, the film is dendritic¹⁹ (Fig. 2). However, above a critical current density, the deposit is uniform, and exhibits a stable flat front on scales of the order of at least a few hundred micrometres, and up to

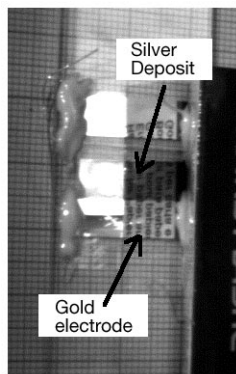


Figure 3 A deposit of silver metal on a cover glass. The cover glass has been metallized over its entire width up to the distance when the current was interrupted. The deposit shows a reflection, indicating that it is of optical quality. $I = 600 \mu\text{A}$, $C = 0.05 \text{ M}$ silver nitrate.

centimetres in width. This critical current density is of the order of 10 mA cm^{-2} for silver, copper and tin.

But even in the regime of covering deposition, the front is destabilized by buoyancy. It is known that gravity currents form in thin cells²¹. These gravity currents have been modelled quantitatively²². The models and the experiments show that the buoyancy effect can be reduced by growing vertically²¹. Therefore, we performed the growth vertically downwards. In this orientation, stability of the fronts can be improved up to two centimetres in width (at present, and especially for silver; Fig. 3).

We also prepared a cylindrical cell composed of a capillary tube 1 mm in diameter in the centre of which we put a glass fibre 100 μm in diameter. The fibre had no conducting coating on it. The fibre was then coated with silver paint at one end only, to serve as the starting point. At the other end of the cell, a silver wire was used as the anode. The solution was 0.05 M silver nitrate. Again, we were able to silver-coat the fibre all around its perimeter in this regime of ultrafast deposition (data not shown); the current passing through the fibre coating was 160 μA , which gives a current density again of the order of 20 mA cm^{-2} (ref. 6).

The process also works well with copper on glass (Fig. 2), and with copper on raw Teflon (Fig. 4), without any gold coating, from either copper sulphate or copper sulphate solution (0.01–0.05 M). We also checked the technique with tin, from tin chloride solutions in the same range of concentrations, and were able to obtain tin films of the same morphology—although the range of currents is more limited, because of hydrogen evolution. Also, unlike copper and silver, in the case of tin an extremely thin film precedes the

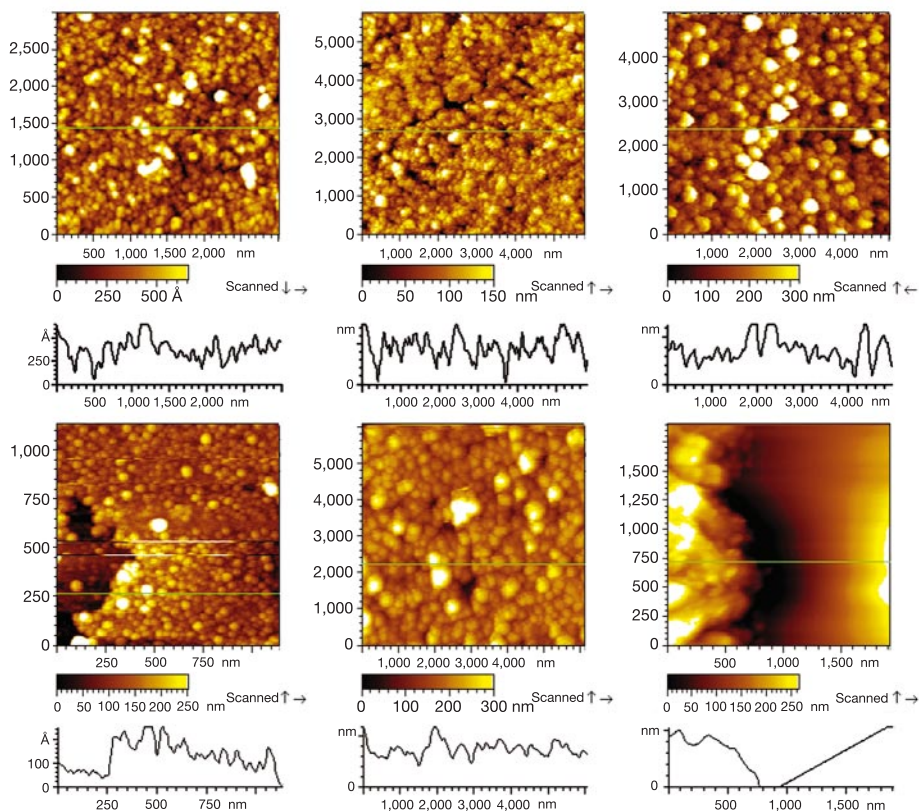


Figure 4 Different samples, as observed by atomic force microscopy *ex situ* (Molecular Imaging, acoustic mode). For each sample, a topographic image (colour) and a horizontal cross-section (black and white graph) is shown. First row, silver deposits. They show a regular covering layer of grains, rather monodisperse in size. Columns 1 and 2: the solution is 0.05 M silver nitrate, currents 1 mA and 0.5 mA. Column 3: same solution but with small amounts of PVA (1 g l^{-1}) as additive, $I = 0.5 \text{ mA}$. Bottom row, columns 1 and 2. Tin films, from tin chloride. The thinnest deposits (column 1), obtained with tin on glass at

lower currents ($I = 0.07 \text{ mA}$) are only 100 $\text{\AA}$ thick. The grains appear to be about 20–50 nm in radius. A hole in the deposit shows that its thickness is spanned by one or just a few grains. The thicker tin deposits (column 2) were obtained in a regime similar to the other metals ($C = 0.04 \text{ M}$, $I = 0.1 \text{ mA}$). Column 3 (bottom right), copper film deposited on Teflon, from copper sulphate solution. In this case, no activation of the surface was done. The copper was deposited directly on the Teflon. The image shows the edge—the deposit is to the left, the Teflon substrate to the right.

formation of the main deposit. At lower current densities, it is possible to deposit only this extremely thin tin film: it is 5 nm thick (Fig. 4), and composed of a carpet of small grains side by side. Whereas the 200-nm copper and 300-nm tin films in Fig. 4 have a thickness close to that predicted by theory, the 5-nm film is much thinner.

We expect that the deposition reported here will be possible with any metal that is known to deposit in the powdery regime of growth, in the shape of rounded crystals. We propose the following mechanistic explanation of this effect. First, in thin cells, and with a binary electrolyte, very high fields are generated at the tips of the deposits¹⁷. These very high fields induce nucleation and growth of a polycrystalline deposit¹⁹. As it is observed that growth is much more rapid in scratches^{5,20}, it is clear that the dangling bonds of glass have catalytic properties, under the action of the large electric field. We now consider why the deposit should be covering for higher currents. As seen in Fig. 2, this surprising stability is not due to an increase in the size of deposit features up to the sample size, but to a progressive closing of voids between ever-smaller branches, in which individual grains become themselves ever smaller¹⁹. This proves that, as the growth speed is increased, the capillary length of individual branches is decreased (as expected from theory^{9,23}). But so is the typical size λ of gradients ahead of the deposit, because¹⁶ $\lambda = (-kT/eE_0)(z_c + z_a)(z_c z_a)(1 + \mu_a/\mu_c)$, where E_0 is the electric field in the bulk, and z_c, z_a, μ_c and μ_a are the charges and mobilities of the cations and anions. When λ becomes smaller than the grain size, instabilities cannot develop, and the front is stable ('absolute stability' in the context of pattern formation). This stability makes it possible to electroplate insulators in conditions very far from equilibrium. The coating of fibres, ribbons and plates seems possible. Many applications of this process may be considered, such as replacing the vapour seeding process in the electronics industry, tailoring mirrors of unusual metals or shapes, and direct coating of organic materials. In general terms, the process proposed here has some advantages over conventional electroless deposition on insulators, in that the film progression and grain size are controlled externally, the process can be interrupted at any time, and it should work with many simple salts, even without additives. But it should be acknowledged that, as the deposition process starts from one end of the sample and progresses towards the other end at speeds of the order of 1 m h^{-1} at most, the overall production output would be much smaller than existing electroless techniques, which coat in approximately 5 minutes glass plates of size $4 \text{ m} \times 4 \text{ m}$.

We note that using this very rapid plating technique with Li, as reported here for Ag, Cu and Sn, might eliminate the cycling problems of Li rechargeable batteries. Indeed, cycling efficiency of Li batteries is drastically reduced by dendritic growth. This is ascribed, in part, to the poor cyclability of a powdery tree. In present designs, dendrites are always seen to grow perpendicularly to the electrodes. A set-up similar to ours would in principle generate a thin film whose morphology is easier to cycle. □

Received 30 April 2001; accepted 25 February 2002.

1. Dini, J. *Electrodeposition* (Noyes Publication, Park Ridge, New Jersey, 1992).
2. Kircher, A. *Mundus Subterraneus, Caput VI, liber duodecimus part I*, Ars Chymurgica (Apud Janssonium et Weyerstraten, Amsterdam, 1664).
3. Kircher, A. *China Monumentis qua Sacris qua Profanis, nec non Variis Naturae et Artis Spectaculis Aliarumque Rerum Meomriabilium Argumentis Illustrata* (Apud Janssonium et Weyerstraten, Amsterdam, 1667).
4. Needham, J. & Peng Yoke, H. *Science and Civilization in China* Vol. 5(2) (Cambridge Univ. Press, Cambridge, 1984).
5. Fleury, V. *Arbres de Pierre, La Croissance Fractale de la Matière* (Flammarion, Paris, 1998).
6. Fleury, V. Patent PCT/FR00/02757 (2000).
7. Bergstrasser, T. R. & Merchant, H. D. in *Defect Structure, Morphology and Properties of Deposits Proceedings of the Materials Week Rosemont 1994* (ed. Merchant, H. D.) 115–168 (Minerals Metals Materials Society, Warrendale, Pennsylvania, 1995).
8. Despic, A. R. & Popov, K. I. in *Modern Aspects of Electrochemistry* Vol. 7 (eds Conway, B. E. & Bockris, J. O'M.) 199–313 (Butterworths, London, 1972).
9. Pelcé, P. *Dynamics of Curved Fronts* (Academic, London, 1991).
10. Meakin, P. *Fractals, Scaling and Growth far from Equilibrium* (Cambridge Univ. Press, Cambridge, 1988).
11. Vicsek, T. *Fractal Growth Phenomena* 2nd edn (World Scientific, Singapore, 1992).

12. Matsushita, M., Sano, M., Hayakawa, Y., Honjo, H. & Sawada, Y. Fractal structure of zinc metal leaves grown by electrodeposition. *Phys. Rev. Lett.* **53**, 286–289 (1984).
13. Fleury, V., Chazalviel, J. N., Rosso, M. & Sapoval, B. The role of the anions in the growth speed of electrochemical deposits. *J. Electroanal. Chem.* **290**, 249–255 (1990).
14. Melrose, J. R., Hibbert, D. B. & Ball, R. C. Interfacial velocity in electrochemical deposition and the Hecker effect. *Phys. Rev. Lett.* **65**, 3009–3012 (1990).
15. Garik, P. *et al.* Laplace and diffusion controlled growth in electrochemical deposition. *Phys. Rev. Lett.* **62**, 2703–2706 (1989).
16. Chazalviel, J. N. Electrochemical aspects of the generation of ramified metallic deposits. *Phys. Rev. A* **42**, 7355–7367 (1990).
17. Fleury, V., Chazalviel, J. N., Rosso, M. & Sapoval, B. Experimental aspects of dense morphology in copper electrodeposition. *Phys. Rev. A* **44**, 6693–6705 (1991).
18. Fleury, V. On a new kind of ramified electrodeposits. *J. Mater. Res. Soc.* **6**, 1169–1174 (1991).
19. Fleury, V. Branched fractal patterns in non-equilibrium electrochemical deposition from oscillatory nucleation and growth. *Nature* **390**, 145–148 (1997).
20. Fleury, V. & Barkey, D. Runaway growth in two-dimensional electrodeposition. *Europhys. Lett.* **36**, 253–258 (1996).
21. Rosso, M., Chazalviel, J. N., Fleury, V. & Chassaing, E. Experimental evidence for gravity induced motion in the vicinity of ramified electrodeposits. *Electrochim. Acta* **39**, 507–515 (1994).
22. Chazalviel, J. N., Rosso, M., Chassaing, E. & Fleury, V. A quantitative study of gravity-induced convection in two-dimensional parallel electrodeposition cells. *J. Electroanal. Chem.* **407**, 61–73 (1996).
23. Ben-Jacob, E. & Garik, P. The formation of patterns in non-equilibrium growth. *Nature* **343**, 523–530 (1990).

Acknowledgements

W.A.W. acknowledges the financial support of Saint-Gobain.

Correspondence and requests for materials should be addressed to V.F. (e-mail: vincent.fleury@polytechnique.fr).

Constraints on radiative forcing and future climate change from observations and climate model ensembles

Reto Knutti, Thomas F. Stocker, Fortunat Joos & Gian-Kasper Plattner

Climate and Environmental Physics, Physics Institute, University of Bern, Silderstr. 5, 3012 Bern, Switzerland

The assessment of uncertainties in global warming projections is often based on expert judgement, because a number of key variables in climate change are poorly quantified. In particular, the sensitivity of climate to changing greenhouse-gas concentrations in the atmosphere and the radiative forcing effects by aerosols are not well constrained, leading to large uncertainties in global warming simulations¹. Here we present a Monte Carlo approach to produce probabilistic climate projections, using a climate model of reduced complexity. The uncertainties in the input parameters and in the model itself are taken into account, and past observations of oceanic and atmospheric warming are used to constrain the range of realistic model responses. We obtain a probability density function for the present-day total radiative forcing, giving 1.4 to 2.4 W m^{-2} for the 5–95 per cent confidence range, narrowing the global-mean indirect aerosol effect to the range of 0 to -1.2 W m^{-2} . Ensemble simulations for two illustrative emission scenarios suggest a 40 per cent probability that global-mean surface temperature increase will exceed the range predicted by the Intergovernmental Panel on Climate Change (IPCC), but only a 5 per cent probability that warming will fall below that range.

The expected future warming of the climate system and its potential consequences increase the need for climate projections with clearly defined uncertainties and likelihood estimates². The IPCC provides these probabilities for most of their findings in the

recently published Third Assessment Report¹. However, some of the most important uncertainties—such as the projected surface warming—are still based on expert judgement, and are only given as ranges derived from different models. The evidence that part of the observed warming of both atmosphere and ocean^{3,4} is caused by anthropogenic emissions of greenhouse gases and aerosols^{1,5–10} may help assess climate models, and has been used to scale model projections for the next few decades^{11,12}. Wigley and Raper¹³ recently presented probabilities for the future warming by performing ensemble simulations with a simplified model calibrated to the same three-dimensional (3D) ocean–atmosphere models as used in the IPCC Third Assessment Report¹. The combination of ensemble simulations that take into account uncertainties in input and model parameters with the use of observational evidence as an independent constraint provides a powerful approach for an objective uncertainty assessment in global warming projections. Here we determine constraints on the climate sensitivity, on the radiative forcing and on the future warming that arise from the requirement that the modelled large-scale surface warming and ocean heat uptake both match observations. We do this by using the reconstructed and projected radiative forcing of all major forcing components in combination with ensemble simulations of a coupled ocean–atmosphere model of reduced complexity (see Methods section).

To illustrate the relationship between radiative forcing, climate sensitivity, ocean mixing and the resulting model response, we have first calculated the global ocean heat uptake (Fig. 1a) and global-

mean surface warming (Fig. 1b) for various set-ups of the ocean model and climate sensitivities. In this study, climate sensitivity is expressed as the increase of global-mean equilibrium surface temperature for a doubling of pre-industrial atmospheric CO₂ concentration. The mean and standard deviation of all model simulations (Fig. 1a and b, solid lines and shaded bands) denote the uncertainty in ocean heat uptake and surface warming due to ocean model uncertainties in mixing properties and surface-to-depth transport of excess heat, an important variable affecting the transient temperature trend¹⁴. This allows us to test recent claims by Barnett *et al.*¹⁰ that observed ocean heat uptake provides the strongest constraint on the climate sensitivity and that climate sensitivity has to be low in order to match the observed ocean heat uptake. However, in their 3D simulations, Barnett *et al.* neglect the radiative effects of changes in solar irradiance, volcanic aerosols and the indirect effect of aerosols, as well as the uncertainties that are attached to the observed ocean heat uptake and surface warming. Taking all these factors into account, we find that climate sensitivity is only weakly constrained by the observed ocean heat uptake.

The comparison of the modelled ocean heat uptake with the instrumental estimates (Fig. 1a, horizontal solid lines) and their uncertainties (one standard deviation, horizontal dotted lines)⁴ yields an approximate climate sensitivity of 5.7 K, with an uncertainty of ± 1 K attributable to model uncertainties and ± 2 K due to uncertainties in the data. The same diagnostics for the modelled surface temperature increase over the last century and the observed surface warming of 0.6 ± 0.1 K (ref. 2) yields a climate sensitivity of 4.6 K with uncertainties of about -1.5 to $+1.5$ K attributable to model uncertainties, and -1.3 to $+2.6$ K due to uncertainties in the data (Fig. 1b). Additional uncertainties arise from uncertainties in radiative forcing. The global-mean indirect aerosol forcing, for example, is estimated by IPCC¹ to be in the range of 0 to -2 W m⁻² for the year 2000. We have performed simulations where either the indirect aerosol effect (Fig. 1c and d, dash-dotted) or the natural (solar and volcanic) forcings (Fig. 1c and d, dashed)

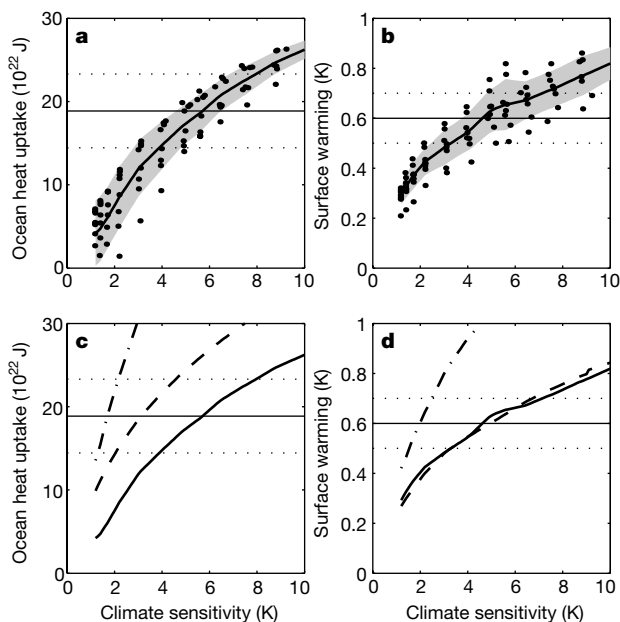


Figure 1 Relation between radiative forcing, climate sensitivity, and modelled atmospheric and oceanic warming. **a, b**, Global ocean heat uptake 1955–95 (to a depth of 3,000 m) and global-mean surface air temperature increase 1900–2000 versus climate sensitivity (expressed as global-mean equilibrium surface temperature increase for a doubling of pre-industrial atmospheric CO₂) for eight model set-ups (different subgrid-scale mixing parameterizations²⁹ and different vertical diffusivities). Calculations were performed using standard reconstructed anthropogenic and natural radiative forcing. Each dot indicates one model simulation. The bold solid curve and shaded band denote the mean and uncertainty (one standard deviation) arising from different ocean mixing properties. Horizontal solid and dotted lines mark the mean and uncertainty (one standard deviation) of the observed ocean heat uptake⁴ and observed surface temperature increase³. **c, d**, Model mean values as in **a** and **b** (solid lines), but when neglecting natural, that is, solar and volcanic, forcings (dashed lines) or when neglecting the indirect aerosol forcing (dash-dotted lines). Constraining the climate sensitivity from the observed warming is mainly hampered by uncertainties in the radiative forcing components and temperature data rather than by the range covered by various set-ups of the climate model used.

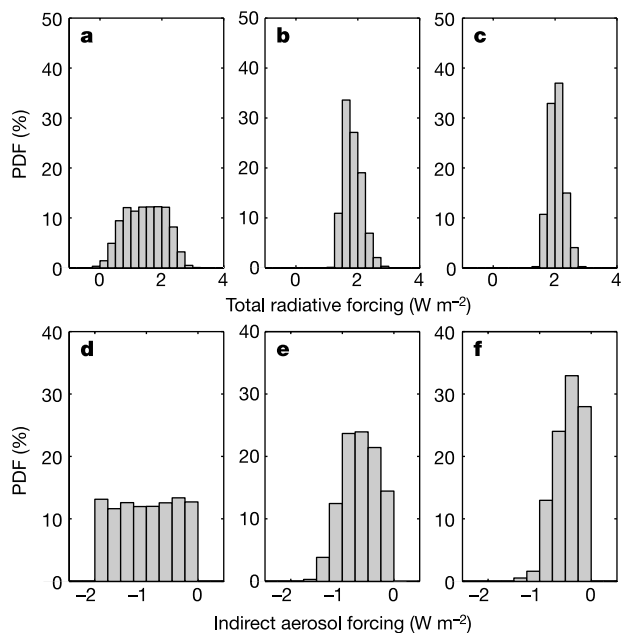


Figure 2 Constraints on the radiative forcing from the observed atmospheric and oceanic warming. Probability density functions (PDF) for the total (anthropogenic and natural) radiative forcing (**a–c**) and the indirect aerosol forcing (**d–f**) in the year 2000 are based on 25,000 Monte Carlo simulations. The initially assumed PDFs are given in **a** and **d**. The requirement that the model matches the temperature observations strongly narrows the PDFs (**b** and **e**). If in addition the climate sensitivity is restricted to the range adopted by the IPCC (1.5–4.5 K), the PDFs in **c** and **f** are obtained.

were neglected. Low estimates for climate sensitivity result from ignoring the indirect aerosol effect, whereas high climate sensitivities are required to match observations when assuming a strong indirect aerosol effect.

From Fig. 1c and d we estimate a climate sensitivity of about 6 K required to match the observed warming trend when applying anthropogenic and natural forcings. Similarly, about 4 K and 2 K are appropriate when neglecting natural forcings or the indirect aerosol effect. The evolution with time of surface warming and ocean heat uptake for these three forcing cases and the standard ocean model set-up shows that the model is able to reasonably reproduce the observed temporal evolution of global-mean surface warming³ for the past 140 years with and without the indirect aerosol forcing (see Supplementary Information). The model has difficulties in reproducing the almost constant temperature between 1940 and 1970 and the strong warming after 1980, indicating that either the assumed radiative forcing is not correct, or part of the observed temperature evolution is due to internal climate variability which is not resolved in this model. When the natural forcing is neglected, much of the variability in the surface warming is lost and

the agreement is significantly worse, consistent with earlier model studies^{6,7,9}. This applies similarly to the modelled ocean heat uptake, except that much less of the decadal structure in the data⁴ is reproduced, consistent with results obtained with 3D models^{10,15}.

To attach probabilities to our results, we have simulated a Monte Carlo set of 25,000 global warming simulations using five ocean model set-ups and taking into account the uncertainties in radiative forcings and climate sensitivity (see Methods section). From those simulations that are consistent with the observed surface warming and ocean heat uptake, we can only derive a very broad probability density function (PDF) for the climate sensitivity (see Supplementary Information) which excludes neither very small climate sensitivities (around 1.2 K, the value if no feedbacks are present) nor unreasonably large values far above the widely accepted IPCC maximum of 4.5 K. This result is in agreement with recent PDF estimates for climate sensitivity, based on the observed surface warming, natural variability and either ocean models^{16,17} or observed ocean warming^{18,19}. This indicates that given the uncertainties in the radiative forcing, in the temperature records, and in currently used ocean models, it is impossible at this stage to strongly constrain the climate sensitivity, as proposed by Barnett *et al.*¹⁰. However, we can strongly constrain the sum of the radiative forcing and thereby the indirect aerosol forcing, the most uncertain of the individual forcing components. If we demand consistency with the temperature records, the PDF of the total radiative forcing for the year 2000 is considerably narrowed (1.4–2.4 W m⁻² for the 5–95% confidence range, Fig. 2b), compared to the initially assumed PDF (Fig. 2a). Furthermore, all results from comprehensive 3D climate models suggest a range for the climate sensitivity of about 1.5 to 4.5 K (ref. 1). By adopting this range as an additional constraint, which is completely independent of this study, a narrower forcing range results (1.6–2.5 W m⁻², Fig. 2c). Assuming the PDFs for the other forcing components to be correct estimates, the procedure reduces the initially assumed uniform PDF for the indirect aerosol forcing (Fig. 2d) considerably. Our analysis suggests that the negative indirect aerosol forcing plus any forcing not explicitly considered for the year 2000 is smaller than 1.2 W m⁻² in magnitude with a probability of 95% (Fig. 2e) for any climate sensitivity, and over 99% if the climate sensitivity is restricted to 1.5 K–4.5 K (Fig. 2f). Even for much less restrictive assumptions regarding the forcing PDFs, these probabilities decrease only slightly to 85% and 95%, respectively (for example, when all IPCC uncertainties are taken as one standard deviation and/or when using a broader PDF for the indirect aerosol forcing). Further simulations suggest that a slightly positive indirect aerosol forcing (plus any forcing not considered) at year 2000 cannot be excluded by this method, if the corresponding PDF assumption is extended to positive values. But although the uncertainty about its magnitude is large, there is general agreement that the indirect aerosol forcing is indeed negative¹.

A key question is whether the observed warming of the twentieth century does constrain the uncertainties in projections of future warming. We investigated the range of projected surface warming until the year 2100 for two illustrative emission scenarios of the IPCC, B1 and A2 (ref. 20), using only those simulations that are consistent with the temperature observations (Fig. 3a). If the observed surface warming and ocean heat uptake are taken as the only constraints, we find a surface temperature increase of 1.6 to 3.8 K (5–95% range, relative to the 1961–90 period) by the year 2100 for scenario B1 (Fig. 3b), with a probability of 40% that the warming exceeds the range given by the IPCC and <5% for a warming below that range. For a restricted range of climate sensitivities (1.5–4.5 K), we estimate 5–95% ranges for the projected warming of 1.5–2.6 K for scenario B1 (Fig. 3c), and 2.5–4.3 K (Fig. 3d) for A2, consistent with the results proposed by the IPCC for these scenarios and based on comprehensive 3D models.

These results are remarkable for several reasons. First, there is no

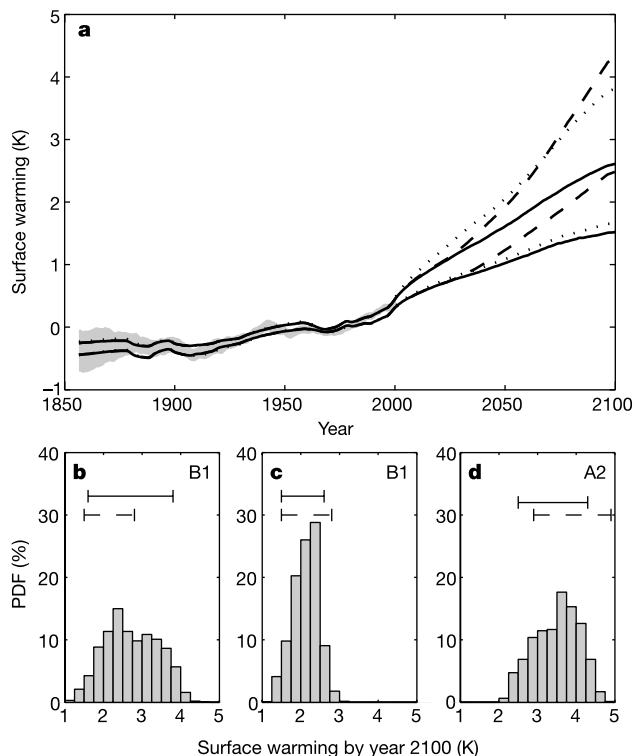


Figure 3 Probabilistic surface warming projections for two IPCC scenarios, as constrained by the observed atmospheric and oceanic warming. **a**, Range (5–95%) of projected global-mean surface air temperature increase (10-yr running mean) for the IPCC SRES illustrative emission scenarios B1 (solid lines, dotted lines) and A2 (dashed lines). Results are obtained from those ensemble members that match observed surface warming (grey shaded band) and observed ocean heat uptake (not shown), for climate sensitivities of 1–10 K (dotted lines for scenario B1) or for the IPCC range 1.5–4.5 K (solid lines for scenario B1, dashed lines for scenario A2). Panels **b–d** show the corresponding PDFs for the surface warming by year 2100. Confidence ranges (5–95%) based on the PDFs, and the uncertainties proposed by IPCC¹, are given as solid and dashed horizontal bars, respectively. When the observed warming is taken as the only constraint, the projected range for scenario B1 largely exceeds the uncertainties given by the IPCC (**b**, and dotted lines in **a**), indicating that currently accepted projections strongly rely on model-derived climate sensitivities and might underestimate the probability of a strong warming. If the range of climate sensitivity is independently constrained (1.5–4.5 K), our 5–95% estimates roughly agree with the uncertainty proposed by the IPCC for scenario B1 (**c**, and solid lines in **a**), and for A2 (**d**, and dashed lines in **a**).

tuning of the simplified model towards more complex models. Thus, our results support the use of simplified models to assess the probabilities and uncertainties of the projections of at least global properties obtained by more complex models. Second, the IPCC results are obtained using different models but a single radiative forcing evolution for a certain scenario. Here we circumvent this limitation, and consider uncertainties in climate sensitivities, ocean mixing, and the reconstructed or projected radiative forcing. Third, taking the observed warming of ocean and atmosphere as the only constraint, we find that the currently accepted IPCC consensus about the uncertainty in global warming projections might significantly underestimate the probability of a strong warming. Our results are in agreement with the IPCC uncertainty ranges only when using a model-derived upper limit of about 5 K as an independent constraint for climate sensitivity. Whereas the IPCC attaches no statistical information to its projection uncertainties, the ensemble approach presented here suggests that the probability distribution of the temperature increase for a certain scenario is similar to a gaussian distribution, whose 5–95% range corresponds to the uncertainties given by the IPCC. The consistency with the observed warming is a strong constraint in our ensemble simulations for at least the next few decades, when the forcing is the main uncertainty. By the end of the century, the projection uncertainty for a particular scenario becomes increasingly dominated by the uncertainty in climate sensitivity.

The simplicity of our model prevents us from taking into account any internal variability of the climate system. We might therefore miss some nonlinear feedbacks, but we benefit from the fact that there is no noise superimposed on the model response. Another powerful method of scaling future climate projections is to use an optimal fingerprint method^{11,12}. This method considers natural variability and does not require an estimate for climate sensitivity, but cannot take into account model and input uncertainties in a probabilistic way. Furthermore, the necessary assumption that the current balance of greenhouse warming and sulphate cooling remains approximately constant makes that method particularly useful for the near future, but not for long-term projections where sulphate forcing is expected to decrease substantially. The temperature ranges derived by this fingerprint method¹¹ for a similar scenario (IS92a) are consistent with our results, from which we conclude that including natural variability would not strongly widen the PDF for long-term temperature changes. We deliberately refrain from an overall probabilistic estimate by combining results from different emission scenarios¹³, as that would depend on a subjective estimate of the likelihood of individual scenarios².

The combination of observed surface air temperature increase and ocean heat uptake with results from ensemble climate simulations provides a strategy to estimate more objectively uncertainties of climate projections. It is desirable that the present results be assessed by ensemble simulations with more comprehensive climate models. Further progress will depend on continuing high-quality observations with global coverage, in particular ocean temperatures, a refined understanding of the climate system, and significantly increased computational resources. □

Methods

The applied climate model consists of a zonally averaged dynamical ocean model, coupled to a zonally and vertically averaged energy- and moisture-balance model of the atmosphere^{21,22}. For efficiency, we use the annual-mean model version, but differences from the results obtained with the seasonal version²³ are negligible. Whereas the climate sensitivity of comprehensive models is determined by the strength of the resolved feedback mechanisms, we specify the radiative perturbation at the tropopause as $\Delta F(t) = \Delta F_{\text{RF}}(t) + \lambda \Delta T_{\text{s}}(t)$, where ΔF_{RF} is the radiative forcing¹. Feedback processes are parameterized in terms of the global-mean surface temperature increase ΔT_{s} , and the constant factor λ is prescribed to obtain different climate sensitivities²⁴. We diagnose the climate sensitivity after 3,000 years of integration. The time history of radiative forcing is prescribed from changes in well-mixed greenhouse gases (CO_2 , CH_4 , N_2O , SF_6 and 28 halocarbons including those controlled by the Montreal Protocol), stratospheric and tropospheric O_3 , the direct forcing of black and organic carbon and sulphate, stratospheric

H_2O owing to CH_4 changes, and the indirect effects of aerosols, all based on simplified expressions that are summarized in refs 1 and 25. Anthropogenic radiative forcing is prescribed from reconstructions for the time 1765–2000, follows a SRES scenario²⁰ from 2000 to 2100, and is kept constant thereafter. For the simulations in Fig. 1, a standard value of -0.8 W m^{-2} is assumed for the indirect aerosol forcing at year 2000, as in earlier studies²⁶. Radiative forcing by volcanoes and variations in solar irradiance are prescribed for the historical period²⁷. Albedo changes due to land use, radiative forcing by dust and the uncertainty in converting future greenhouse-gas emissions into concentrations²⁵ are not considered.

For the Monte Carlo simulations, we have calculated 25,000 global warming simulations using five ocean model set-ups and taking into account the uncertainties in radiative forcings and climate sensitivity. For each forcing component of every individual simulation, a random number (representing the radiative forcing for year 2000) is determined, to which the time history and future projection of that forcing component is scaled. These random numbers are chosen in such a way that their distribution follows the prescribed PDF of the forcing for the year 2000. A gaussian PDF is assumed where absolute uncertainties are given, a log-normal PDF where the uncertainty is expressed as a factor. We assume the uncertainties given by IPCC to be two standard deviations, although the IPCC attaches no statistical meaning to them. For the indirect aerosol forcing, the probability is assumed to be uniform between 0 and -2 W m^{-2} (see refs 25 and 28 for details about the radiative forcing assumptions). For each simulation, a climate sensitivity is randomly chosen between 1 and 10 K (uniform PDF).

An individual model simulation is considered to be consistent with observations if the simulated differences in both global-mean surface temperature between 1900 and 2000 and ocean heat content between 1955 and 1995 match observations within their uncertainties (two standard deviations) and if a correlation criterion indicates reasonable time-dependent agreement for both surface warming and ocean heat uptake.

Specifically, we divide the difference of the observed and modelled warming by the uncertainty of the observed warming (thereby expressing the model mismatch in terms of observation uncertainties), average over time and prescribe a maximum value for this quantity. The main conclusions do not depend on the exact choice of the consistency criteria. However, we implicitly assume that the long-term trends in observed surface warming and ocean heat uptake are due to natural and anthropogenic forcings, and that internal variability only contributes to decadal changes. Further, the results of this study are only weakly sensitive to the assumed forcing PDFs and ocean mixing properties, as unrealistic input/model combinations due to less restrictive input assumptions are usually eliminated by the observational constraints. For example, the indicated limits of surface temperature ranges at year 2100 change by less than $\pm 0.2 \text{ K}$ when assuming different ocean model versions or less restrictive PDFs of the radiative forcing.

Received 27 December 2001; accepted 22 February 2002.

- Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
- Schneider, S. H. What is 'dangerous' in climate change? *Nature* **411**, 17–19 (2001).
- Jones, P. D., New, M., Parker, D. E., Martin, S. & Rigor, I. G. Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **37**, 173–199 (1999).
- Levitus, S., Antonov, J. I., Boyer, T. P. & Stephens, C. Warming of the world ocean. *Science* **287**, 2225–2229 (2000).
- Santer, B. D. *et al.* A search for human influences on the thermal structure of the atmosphere. *Nature* **382**, 39–46 (1996).
- Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. & Mitchell, J. F. B. Causes of twentieth-century temperature change near the Earth's surface. *Nature* **399**, 569–572 (1999).
- Stott, P. A. *et al.* External control of 20th century temperature by natural and anthropogenic forcing. *Science* **290**, 2133–2137 (2000).
- Hegerl, G. C. *et al.* Optimal detection and attribution of climate change: sensitivity of results to climate model differences. *Clim. Dyn.* **16**, 737–754 (2000).
- Stott, P. A. *et al.* Attribution of twentieth century temperature change to natural and anthropogenic causes. *Clim. Dyn.* **17**, 1–21 (2001).
- Barnett, T. R., Pierce, D. W. & Schnur, R. Detection of anthropogenic climate changes in the world's oceans. *Science* **292**, 270–274 (2001).
- Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. & Delworth, T. L. Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* **407**, 617–620 (2000).
- Allen, M. R. *et al.* Quantifying anthropogenic influence on recent near-surface temperature change. *Surv. Geophys.* (in the press).
- Wigley, T. M. L. & Raper, S. C. B. Interpretation of high projections for global-mean warming. *Science* **293**, 451–454 (2001).
- Hansen, J., Russell, G., Laci, A., Fung, I. & Rind, D. Climate response times: Dependence on climate sensitivity and ocean mixing. *Science* **229**, 857–859 (1985).
- Levitus, S. *et al.* Anthropogenic warming of Earth's climate system. *Science* **292**, 267–270 (2001).
- Forest, C. E., Allen, M. R., Stone, P. H. & Sokolov, A. P. Constraining uncertainties in climate models using climate change detection techniques. *Geophys. Res. Lett.* **27**, 469–472 (2000).
- Andronova, N. & Schlesinger, M. E. Objective estimation of the probability distribution for climate sensitivity. *J. Geophys. Res.* **106**, 22605–22612 (2001).
- Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A. & Rayner, N. A. An observationally based estimate of the climate sensitivity. *J. Clim.* (submitted).
- Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. & Webster, M. D. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117 (2002).
- Nakićenović, N. *et al.* *Special Report on Emission Scenarios* (Intergovernmental Panel on Climate Change, Cambridge Univ. Press, 2000).
- Stocker, T. F., Wright, D. G. & Mysak, L. A. A zonally averaged, coupled ocean-atmosphere model for paleoclimate studies. *J. Clim.* **5**, 773–797 (1992).
- Schmittner, A. & Stocker, T. F. The stability of the thermohaline circulation in global warming experiments. *J. Clim.* **12**, 1117–1133 (1999).
- Schmittner, A. & Stocker, T. F. A seasonally forced ocean-atmosphere model for paleoclimate studies.

- J. Clim.* **14**, 1055–1068 (2001).
24. Plattner, G.-K., Joos, F., Stocker, T. F. & Marchal, O. Feedback mechanisms and sensitivities of ocean carbon uptake under global warming. *Tellus B* **53**, 564–592 (2001).
25. Joos, F. *et al.* Global warming feedbacks on terrestrial carbon uptake under the IPCC emission scenarios. *Glob. Biogeochem. Cycles* **15**, 891–907 (2001).
26. Kattenberg, A. *et al.* in *IPCC Second Scientific Assessment of Climate Change* (eds Houghton, J. T. *et al.*) 285–357 (Cambridge Univ. Press, Cambridge, 1996).
27. Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
28. Boucher, O. & Haywood, J. On summing the components of radiative forcing of climate change. *Clim. Dyn.* **18**, 297–302 (2001).
29. Knutti, R., Stocker, T. F. & Wright, D. G. The effects of subgrid-scale parameterizations in a zonally averaged ocean model. *J. Phys. Oceanogr.* **30**, 2738–2752 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank T. Crowley for providing the volcanic and solar radiative forcing data. This work was supported by the Swiss National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.F.S. (e-mail: stocker@climate.unibe.ch).

Origins and estimates of uncertainty in predictions of twenty-first century temperature rise

Peter A. Stott* & J. A. Kettleborough†

* Hadley Centre for Climate Prediction and Research, Met Office, Bracknell, Berkshire RG12 2SY, UK

† Space Science and Technology Department, Rutherford Appleton Laboratory, Didcot, Oxfordshire OX11 0QX, UK

Predictions of temperature rise over the twenty-first century are necessarily uncertain, both because the sensitivity of the climate system to changing atmospheric greenhouse-gas concentrations, as well as the rate of ocean heat uptake, is poorly quantified^{1,2} and because future influences on climate—of anthropogenic as well as natural origin—are difficult to predict³. Past observations have been used to help constrain the range of uncertainties in future warming rates, but under the assumption of a particular scenario of future emissions⁴. Here we investigate the relative importance of the uncertainty in climate response to a particular emissions scenario versus the uncertainty caused by the differences between future emissions scenarios for our estimates of future change. We present probabilistic forecasts of global-mean temperatures for four representative scenarios for future emissions⁵, obtained with a comprehensive climate model. We find that, in the absence of policies to mitigate climate change, global-mean temperature rise is insensitive to the differences in the emissions scenarios over the next four decades. We also show that in the future, as the signal of climate change emerges further, the predictions will become better constrained.

An estimate of the uncertainty in a climate-model-based prediction of twenty-first century global-mean temperature rise is a potentially valuable tool for policy makers and planners^{3,6,7}. Large and difficult-to-quantify uncertainties surround predictions of future demographic changes, economic development and technological change, which will determine future anthropogenic emis-

sions of greenhouse gases and other pollutants. Even with perfect knowledge of emissions, uncertainties in the representation of atmospheric and oceanic processes by climate models limit the accuracy of any estimate of the climate response. Natural variability, generated both internally and from external forcings such as changes in solar output and explosive volcanic eruptions, also contributes to the uncertainty in climate forecasts.

Recently a technique has been developed to quantify uncertainty in predictions by comparing simulations of past temperature changes with observations⁴. Under this approach, based on those developed for the detection and attribution of climate change^{8,9}, we estimate the factors (with associated uncertainties) by which the model's simulated response to various external forcings over the twentieth century can be scaled up or down while remaining in agreement with the observations. The most important external forcings are well-mixed greenhouse gases, other anthropogenic pollutants such as sulphate aerosols (which are produced by oxidation of sulphur dioxide), changes in tropospheric ozone (which is controlled by photochemical reactions), stratospheric ozone depletion, and natural external forcings such as variations in solar irradiance and stratospheric aerosol from volcanic eruptions.

Temperatures will fluctuate about their mean climatic state owing to natural internal variability. We include decadal variability in our uncertainty analysis, but we do not consider sub-decadal variations that would be additional to the uncertainty in decadal temperatures presented here. We also consider fluctuations due to potential future changes in solar output and volcanic eruptions. As it is not possible to predict deterministically changes in natural forcings, we estimate natural external variability from simulations of the past 140 years that include these natural forcings.

The IPCC, in their Special Report on Emissions Scenarios (SRES⁵), has developed a wide range of future emissions scenarios, based on a variety of narrative 'storylines', each describing a possible future development of population, economies and energy sources. The range of scenarios includes interventions leading to reductions in sulphur emissions and introduction of new energy technologies, but does not include additional initiatives to mitigate climate change. Any estimate of socio-economic trends over the course of the twenty-first century is necessarily very uncertain and highly subjective. Our interest here lies in determining the range of likely future climates consistent with current observations under a representative range of emissions scenarios, and investigating how this uncertainty range will change as the signal of climate change becomes stronger.

We can address these questions with predictions of a coupled atmosphere–ocean general circulation model (AOGCM) using a representative subset of emissions scenarios which span most of the total SRES range, without assigning relative probabilities to the different emissions scenarios. On the assumption that a model that over- or under-estimates the climate response by a certain fraction now will continue to over- or under-estimate it by a similar fraction in the future, we can use a comparison between simulated and observed changes over the past 100 years to calculate the uncertainty in a prediction (according to a particular emissions scenario) over the next 100 years. This assumption appears to be justified for global-mean temperature by the AOGCM runs currently available, all of which evolve similarly over time in response to a given forcing despite differences in sensitivity and thus response amplitude¹⁰. Even though climate sensitivity is not well constrained by the observed temperature record^{4,11}, perturbation analysis of simple⁴ and intermediate-complexity¹¹ models indicates that there is a generally linear relationship between past and future global temperature change as we vary the sensitivity of a climate model that continues to hold for unmitigated forcing until the end of the century¹², provided the climate sensitivity and sulphate aerosol forcing are not outside the likely range estimated by the IPCC

Third Assessment Report¹³. This relationship depends on the absence of such possible feedbacks as a shutdown of the thermohaline circulation¹⁴ or the land biosphere switching from being a weak sink for carbon to a strong source¹⁵, and is therefore more likely to hold early on in the century when forcing changes are smallest.

For scenarios in which forcing increases are not sustained, such as stabilization scenarios, an alternative is to calculate probability density functions (PDFs) of key properties of the climate system by varying uncertain model properties and forcings in a simplified climate model^{1,2}. This complementary approach produces observational constraints on climate sensitivity and net aerosol forcing but relies on an idealized model.

The model we use in this analysis is HadCM3^{16,17}, an AOGCM that does not use flux adjustment. To constrain the effects of model uncertainty on future predictions, we have made three ensembles, each of four simulations, of the period 1860–2000. All four simulations in each ensemble are identical except for their initial conditions, which were taken from model states 100 years apart in a 1,820-year control run of HadCM3 in which external forcings have no year-to-year variations. The simulations in the first ensemble incorporate changes in individual well-mixed greenhouse gases including carbon dioxide and methane. Simulations in the second ensemble include, in addition to changes in well-mixed greenhouse gases, changes in tropospheric and stratospheric ozone, and changes in sulphur emissions. The direct effect of sulphate aerosols on planetary albedo is simulated using a fully interactive sulphur cycle scheme that models the emission, transport, oxidation and removal of sulphur species. The indirect effect of tropospheric aerosol, by way of cloud reflectivity¹⁸, is also represented in the model. The simulations in the third ensemble include only natural forcings due to changes in the amount of stratospheric aerosols following volcanic eruptions¹⁹ and spectrally resolved changes in solar irradiance²⁰.

We have also made a series of model predictions for the twenty-first century²¹. For these predictions, we have chosen four representative SRES marker scenarios (A1FI, A2, B1 and B2), one each from the four SRES storylines and scenario families, and each of which assumes a distinctly different direction for future developments⁵. For each marker scenario we have made two sets of runs starting in 1989. In the first, all anthropogenic forcings are included, whereas the second includes only well-mixed greenhouse gases. As the climate response to greenhouse gases is subject to

different sources of uncertainty from those due to sulphate aerosols, we have included the two effects separately in our analysis to reflect the fact that they may be subject to different errors. We have available three predictions, including all anthropogenic forcings according to the A2 scenario, which we average, and two simulations according to the B2 scenario, which we also average. For the other two scenarios we have only a single simulation including all anthropogenic forcings, and for each of the four scenarios we have a single simulation which includes changes in greenhouse gases only.

Uncertainty ranges of global-mean temperature rise for our four representative emissions scenarios are shown in Fig. 1. (Further details of the methodology are given in the Methods section.) In the first three decades of the twenty-first century, they overlap considerably and indicate that the 5 to 95 percentiles of the forecast temperature distribution are 0.9 to 1.9 K relative to pre-industrial (control) climate by the 2020–30 decade. Only after 2050 is the best estimate of the scenario prediction that warms the greatest over the century (the fossil fuel intensive scenario A1FI) outside the 90% confidence interval of the prediction that warms the least, B1. In the latter half of the century, all SRES scenarios have much reduced SO₂ emissions increasing the uncertainty range consistent with past observations⁴. By the end of the century temperatures are predicted to rise from pre-industrial values by 1.9 to 4.0 K according to the B1 scenarios and by 3.6 to 7.5 K according to the A1FI scenario.

Expressed relative to 1990–2000 (Fig. 2), temperatures are predicted to rise between 0.3 and 1.3 K (5–95%) by 2020–30 whichever scenario we take, and by the end of the century by 1.2 to 3.3 K according to the B1 scenario and by 3.0 and 6.9 K according to the A1FI scenario. The upper uncertainty limits for temperature change by 2100 of less than 5 K (1.7–4.9 K, ref. 6; 1.1–4.5 K, ref. 7) depend on assigning a low probability to the A1FI scenario in the absence of climate mitigation policies. The relative contributions to the total forecast uncertainty from externally generated natural variability are greatest early in the century. If we consider anthropogenic warming alone⁴, temperatures are predicted to rise from their 1990–2000 values between 0.4 and 1.2 K by 2020–30. The dominant contributor to uncertainty in this decade is that due to response uncertainty.

As the signal of climate change becomes stronger, future predictions of climate change become better constrained. We estimate that by 2020, uncertainties in late twenty-first century global warming

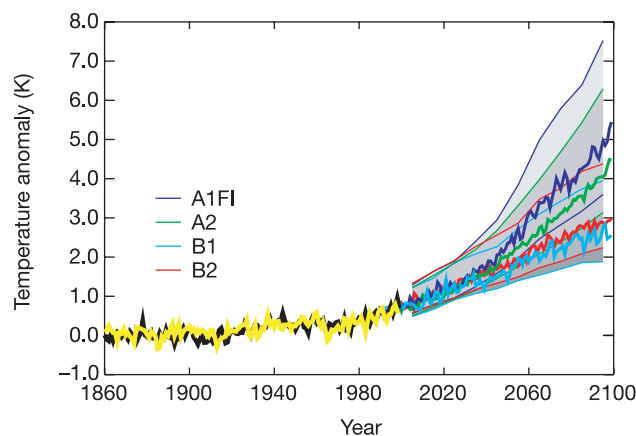


Figure 1 Uncertainty in predicted global-mean near-surface temperatures. Shown are predictions according to the SRES A1FI (blue), A2 (green), B1 (light blue) and B2 (red) scenarios, with their uncertainties (5 to 95 percentiles) shown as grey regions bounded by lines of the appropriate colour. Also shown are the observed temperatures in black, and a

HadCM3 simulation including both anthropogenic (greenhouse gases, sulphate aerosols and tropospheric and stratospheric ozone) and natural (solar and volcanic) forcings in yellow.

could potentially reduce to half their values today (Fig. 2). By taking a prediction that follows the B2 scenario to represent a possible future world over the next 20 years and using temperature change observed over the 1920–2019 period to constrain predictions from 2020, we find that global-mean temperatures would be predicted to rise by the end of the century by 2.3 to 3.3 K (5–95%) according to the B1 scenario and by 4.5 to 6.3 K according to the A1FI scenario, relative to pre-industrial values. This calculation assumes a perfect knowledge, to within natural internal variability, of the climatic effects of external forcings from 2000 to 2019, which provides an optimistic estimate of the potential reduction in uncertainty, although diagnostics other than surface temperature might help to reduce uncertainty more rapidly. Also, the spread between the PDFs is likely to change as more information becomes available over the next 20 years about the likely range of emissions paths over the remainder of the twenty-first century.

Our global-mean temperature forecasts, with their associated uncertainty ranges, are based on an objective comparison of the model's simulation of past temperatures with observations and do not depend on a subjective estimate of the model's climate sensitivity based on "expert elicitation"⁶. They are also based on a comprehensive climate model that accurately simulates multi-decadal variations in global-mean near-surface temperatures over the twentieth century when it includes the most important anthropogenic and natural forcings²². They show that by the 2020–30 decade, global-mean near-surface temperatures are likely to be 0.3 to 1.3 K warmer than they were three decades earlier, regardless of whether CO₂ emissions increase only by 1 Gt C yr⁻¹ from their 1990 levels, as the B1 scenario predicts, or by 8 Gt C yr⁻¹ as A1FI predicts¹³. This lack of sensitivity to increases in CO₂ emissions arises partly because the climate system's inertia dictates that much of the rise is already committed by pre-2000 emissions, and partly because differences in CO₂ emissions are offset to some extent in most scenarios by differences in SO₂ emissions.

In our analysis, we account for errors in the magnitude of forecast patterns of temperature change but do not account for errors in the patterns themselves; consequently our approach may not be valid for regional or non-temperature indicators of climate change. To explore fully the influence of model error requires a systematic

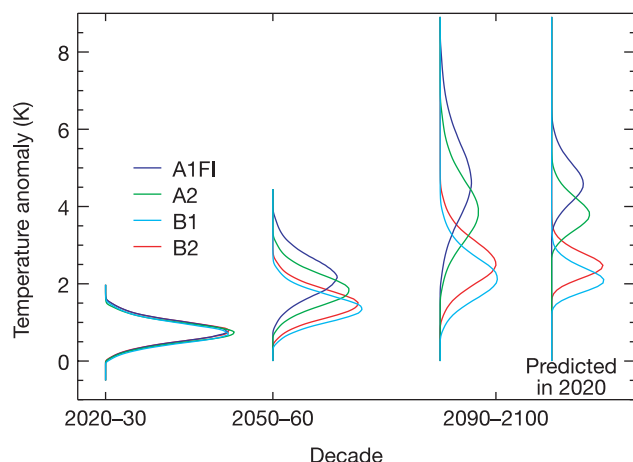


Figure 2 Probability density functions (PDFs) of temperature change. Shown are PDFs for four SRES scenarios (A1FI, A2, B1 and B2) for 2020–30, 2050–60 and 2090–2100 decades relative to the 1990–2000 decade, calculated by constraining HadCM3 simulations to the observed temperature change over the 1900–99 period. The PDFs at the far right are for the 2090–2100 decade calculated by constraining HadCM3 simulations to be consistent with the observed temperature change over the 1920–2019 period, where the observations are assumed to follow a B2 scenario prediction after 1999.

evaluation of the sensitivity of coupled AOGCM predictions to varying the many model parameters across their ranges of plausible values²³. Until such an experiment is performed, a linear analysis represents the most objective approach available for predicting future global-mean temperatures using comprehensive models.

In the first four decades of this century, Fig. 1 shows that predictions for a representative ranges of SRES scenarios differ remarkably little. Until 2040, uncertainty is dominated not by uncertainties in emissions scenarios but by uncertainties in climate response. Later in the century, the effect of different emissions matter much more, although even at the end of the century climate response uncertainty remains as great as emissions uncertainty. Relatively modest increases in global temperature by 2100 are possible under some emissions scenarios, but extremely rapid warming rates over the century cannot be excluded under fossil fuel intensive scenarios. □

Methods

The method we use to determine the uncertainty range for each forecast scenario is as follows. First we compare the observed evolution of the surface temperature over the twentieth century with the modelled response to greenhouse gases, sulphate aerosols and natural forcing. This generates a distribution of factors by which the model response to each forcing can be scaled up or down while still matching the observations. Second, we account for uncertainty in the model forecasts, which would be present even if we were certain about the past response to forcing. Third, we combine the distributions of the scaling factors with the distributions of the model forecast to generate a distribution of the forecast temperature response to the combined forcing. Each of these stages is outlined below.

Calculation of scaling factors

To compare model simulations with observations and calculate the scaling factors to be applied to model predictions of temperature changes due to greenhouse gases and other anthropogenic forcings, we follow the procedure of ref. 24. Decadal-mean data from models and observations are smoothed spatially to retain only scales greater than 5,000 km (ref. 25). Scaling factors on the greenhouse gases, and other anthropogenic and natural contributions are estimated using standard optimal fingerprinting²⁶, modified to take account of sampling noise in model-simulated signals²⁷. Intra-ensemble differences are used to define the optimization space, and a 1,820-year control run of HadCM3 is used for uncertainty analysis. The optimal fingerprinting gives a distribution of scaling factors for the forced response. The uncertainty implied by the distributions results from interdecadal internal variability, which prevents us determining with certainty how much of twentieth-century temperature variation was externally forced and how much can be accounted for by internal variability.

Uncertainty in model forecasts

Even if there was no uncertainty in the past response of the climate system to external forcing, there are sources of uncertainty in the forecast made by a model. The first source of uncertainty is due to uncertainty in future natural forcing. Here we take this into account by assuming there is no deterministic knowledge of natural forcings over future decades. We calculate the variance in global-mean temperatures due to changes in solar output and volcanic eruptions from the natural ensemble of four simulations described earlier, adjusting the variance to take account of the finite ensemble size by subtracting one-quarter of the variance due to internal variability of the control run. A distribution of possible forecasts of natural climate change is then generated by selecting random members of a normal distribution whose variance is that of this estimate of natural externally forced variability. Note that the use of a normal distribution will probably underestimate the contribution of low-probability, high-impact natural events, but we are focusing here on the central 5–95% uncertainty ranges.

The second source of uncertainty in the model forecast results from internal variability and from having only a small number of predictions from forecast ensembles. We take account of this uncertainty by selecting random perturbations from a normal distribution and adding these to the anthropogenic predictions. The variance of the normally distributed perturbations is calculated from the control run, and adjusted to account for the size of the forecast ensemble.

Combination of scaling factors with model forecasts

Having obtained distributions of future anthropogenic and naturally forced responses, we combine these with the distribution of factors by which the modelled response must be scaled to match observations of 1900–2000 and additionally include a sample of the internal variability on the top of the forced response. The result is a probability distribution of forecast temperature for each scenario, from which the median temperature forecast and confidence intervals can be determined. In the calculation of temperature changes relative to 1990–2000, we additionally take account of the extra uncertainty arising from taking the difference between two uncertainly known decadal-mean temperatures by doubling the variance due to internal variability about the mean state.

Our uncertainty limits depend on model-based estimates of both internally generated and externally generated natural variability. Reconstructions of past temperature change based on palaeodata indicate that coupled models' internal variability is generally

consistent with that observed²⁸, and HadCM3, when it includes both anthropogenic and natural forcings, simulates many features of observed twentieth-century temperature change²², indicating some success in incorporating external forcings including those due to solar changes and volcanic aerosol. Nevertheless, the current dependence on model-based rather than observationally based estimates of natural variability needs to be tested further against observational evidence. We do not include the effect of observational error in our analysis. The effect of observational sampling error on detection and attribution results has been shown to be small²⁹, but we do not as yet have an estimate of the effects of systematic instrumental errors, such as changes in measurement practices or urbanization.

Received 6 November 2001; accepted 27 February 2002.

1. Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. & Webster, M. Quantifying uncertainties in climate system properties using recent climate observations. *Science* **295**, 113–117 (2002).
2. Knutti, R., Stocker, T. F., Joos, F. & Plattner, G.-K. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723 (2002).
3. Allen, M. R., Raper, S. C. B. & Mitchell, J. F. B. Uncertainty in the IPCC's third assessment report. *Science* **293**, 430–433 (2001).
4. Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. & Delworth, T. L. Uncertainty in forecasts of anthropogenic climate change. *Nature* **407**, 617–620 (2000).
5. Nakicenovic, N. & Swart, R. *Special Report on Emission Scenarios* (Cambridge Univ. Press, Cambridge, 2000).
6. Wigley, T. M. L. & Raper, S. C. B. Interpretation of high projections for global-mean warming. *Science* **293**, 451–454 (2001).
7. Reilly, J. *et al.* Uncertainty and climate change assessments. *Science* **293**, 430–433 (2001).
8. Allen, M. R. & Tett, S. F. B. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419–434 (1999).
9. Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. & Mitchell, J. F. B. Causes of twentieth century temperature change near the Earth's surface. *Nature* **399**, 569–572 (1999).
10. Cubasch, U. *et al.* in *Climate Change 2001: The Scientific Basis* Ch. 9 (eds Houghton, J. T. *et al.*) 525–582 (Cambridge Univ. Press, Cambridge, 2001).
11. Forest, C. E., Allen, M. R., Stone, P. H. & Sokolov, A. P. Constraining uncertainties in climate models using climate change detection techniques. *Geophys. Res. Lett.* **27**, 569–572 (2000).
12. Booth, B. B. B., Allen, M. R. & Kettleborough, J. A. *Exploring the Linearity of the Climate Response to External Forcing* (Tech. Report RAL-TR2002-005, Rutherford Appleton Laboratory, Chilton, UK, 2002).
13. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
14. Stocker, T. F. & Schmittner, A. Influence of CO₂ emission rates on the stability of the thermohaline circulation. *Nature* **388**, 862–865 (1997).
15. Cox, P. N., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
16. Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16**, 147–168 (2000).
17. Pope, V. D., Gallani, M. L., Rowntree, P. R. & Stratton, R. A. The impact of new physical parametrizations in the Hadley Centre climate model—HadAM3. *Clim. Dyn.* **16**, 123–146 (2000).
18. Twomey, S. A. Pollution and the planetary albedo. *Atmos. Environ.* **8**, 1251–1256 (1974).
19. Sato, M., Hansen, J. E., McCormick, M. P. & Pollack, J. B. Stratospheric aerosol optical depths (1850–1990). *J. Geophys. Res.* **98**, 22987–22994 (1993).
20. Lean, J., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610: implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
21. Johns, T. C. *et al.* *Anthropogenic Climate Change for 1860 to 2100 Simulated with the HadCM3 Model under Updated Emissions Scenarios* (Tech. Note 22, Hadley Centre for Climate Prediction and Research, Bracknell, UK, 2001).
22. Stott, P. A. *et al.* External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137 (2000).
23. Allen, M. R. Do-it-yourself climate prediction. *Nature* **401**, 642 (1999).
24. Stott, P. A. *et al.* Attribution of twentieth century temperature change to natural and anthropogenic causes. *Clim. Dyn.* **17**, 1–21 (2001).
25. Stott, P. A. & Tett, S. F. B. Scale-dependent detection of climate change. *J. Clim.* **11**, 32892–3294 (1998).
26. Hasselmann, K. Multi-pattern fingerprint method for detection and attribution of climate change. *Clim. Dyn.* **13**, 601–611 (1997).
27. Allen, M. R. & Stott, P. A. Estimating signal amplitudes in optimal fingerprinting. Part I: theory. *Clim. Dyn.* (submitted).
28. Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
29. Hegerl, G. C., Jones, P. D. & Barnett, T. P. Effect of observational sampling error on the detection and attribution of anthropogenic climate change. *J. Clim.* **14**, 198–207 (2001).

Acknowledgements

We thank many colleagues at the Hadley Centre who developed HadCM3, and who made many of the simulations described here; we also thank M. Allen for advice, out of whose work on quantifying uncertainty in forecasts of anthropogenic climate change this work springs. This work was supported by the UK Department of Environment, Food and Rural Affairs.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.A.S. (e-mail: peter.stott@metoffice.com).

Using the fossil record to estimate the age of the last common ancestor of extant primates

Simon Tavaré*, Charles R. Marshall†, Oliver Will‡, Christophe Soligo§|| & Robert D. Martin¶

* Departments of Biological Sciences, Mathematics, and Preventive Medicine, University of Southern California, Los Angeles, California 90089-1340, USA

† Departments of Organismic and Evolutionary Biology, and Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA

‡ Department of Statistics, University of Washington, Seattle, Washington 98195, USA

§ Anthropologisches Institut & Museum, Universität Zürich-Irchel, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

|| Human Origins Group, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

¶ Academic Affairs, Field Museum, Chicago, Illinois 60605-2496, USA

Divergence times estimated from molecular data often considerably predate the earliest known fossil representatives of the groups studied. For the order Primates, molecular data calibrated with various external fossil dates uniformly suggest a mid-Cretaceous divergence from other placental mammals, some 90 million years (Myr) ago^{1–9}, whereas the oldest known fossil primates are from the basal Eocene epoch (54–55 Myr ago). The common ancestor of primates should be earlier than the oldest known fossils^{10,11}, but adequate quantification is needed to interpret possible discrepancies between molecular and palaeontological estimates. Here we present a new statistical method, based on an estimate of species preservation derived from a model of the diversification pattern, that suggests a Cretaceous last common ancestor of primates, approximately 81.5 Myr ago, close to the initial divergence time inferred from molecular data. It also suggests that no more than 7% of all primate species that have ever existed are known from fossils. The approach unites all the available palaeontological methods of timing evolutionary events: the fossil record, extant species and clade diversification models.

Although several molecular studies indicate that the lineage leading to primates diverged from other eutherian mammals about 90 Myr ago, diagnostic morphological features of primates possibly emerged later, potentially explaining why recognizable

Table 1 Relative sampling intensities for the primate fossil record

Epoch	k	T_k	Observed number of species, D_k	Relative sampling intensity, p_k	
				Scheme 1	Scheme 2
Late Pleistocene	1	0.15	19	1.0	1.0
Middle Pleistocene	2	0.9	28	1.0	1.0
Early Pleistocene	3	1.8	22	1.0	1.0
Late Pliocene	4	3.6	47	1.0	1.0
Early Pliocene	5	5.3	11	1.0	0.5
Late Miocene	6	11.2	38	1.0	0.5
Middle Miocene	7	16.4	46	1.0	1.0
Early Miocene	8	23.8	36	1.0	0.5
Late Oligocene	9	28.5	4	1.0	0.1
Early Oligocene	10	33.7	20	1.0	0.5
Late Eocene	11	37.0	32	1.0	1.0
Middle Eocene	12	49.0	103	1.0	1.0
Early Eocene	13	54.8	68	1.0	1.0
Pre-Eocene	14		0	0.1	0.1

Data are shown for a total of 235 modern species. References for the data can be found in the Supplementary Information.

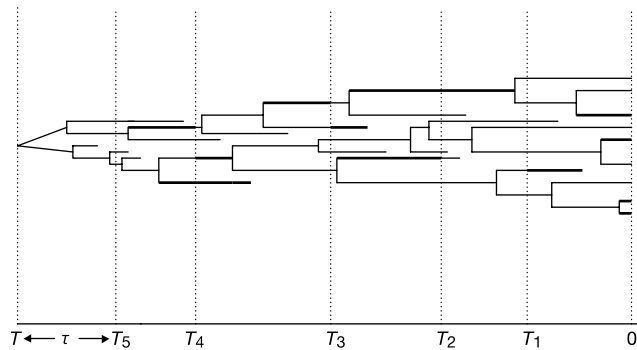


Figure 1 An illustration of the stochastic model of fossil finds. Bases of five stratigraphic intervals at $T_1, \dots, T_5$ Myr ago are shown along the x axis. The temporal gap between the base of the final interval and the point at which the two founding species originate is

denoted by τ . Thick lines indicate species found in the fossil record. Time 0 is the present day.

primates appear late in the known fossil record. However, divergence between strepsirrhines (lemurs and lorises) and haplorhines (tarsiers and anthropoids), postdating development of defining morphological features of primates, probably occurred soon after primates diverged from other mammals^{5,10}. We estimate the divergence time of strepsirrhines and haplorhines—that is, the time of the last common ancestor (LCA) of living primates—using data from the fossil record.

Existing statistical methods designed to account for incompleteness of the fossil record typically use the size and distribution of gaps within observed stratigraphic ranges of lineages to estimate the size of the temporal gap between the oldest fossil and the LCA of the lineages¹². Although these methods are useful for species already known from the fossil record, they are inappropriate for estimating the time of the LCA of higher taxonomic groups because they cannot account for species not preserved at all^{10,12}. Our method for estimating the temporal gap uses an estimate of the proportion of species from the group actually preserved in the fossil record and the shape of its diversification curve. The speciation model we use is the non-homogeneous Markov branching process^{13,14} (Fig. 1).

We have developed a computational approach for estimating the length of the temporal gap between the oldest known fossils and the LCA of a taxonomic group, as well as an estimated confidence interval (see Methods). We use as input the number of extant species, the mean species lifetime, the ages of the bases of the relevant stratigraphic intervals, the numbers of fossil species found in those intervals and the relative sizes of the sampling intensities in each interval. Estimates of the absolute values of the sampling intensities may also be found (see Methods, equation (6)). This gives an estimate of the proportion of species that existed in an interval that were found as fossils.

Different diversification models can be explored with our approach, but we present only results for a logistic diversification model. Logistic growth is the most biologically realistic model¹⁵, matching the general expectation of an equilibrium diversity level. In our case, equilibrium diversity is achieved not through an increase in per-species extinction rate coupled with a decrease in per-species origination rate^{16,17}, but solely through a decrease in origination rate; our extinction rate is independent of standing diversity^{18,19}. We parametrize logistic growth by the time at which diversity reached 90% of its present-day value. In the absence of any simple means of assessing relative sizes of sampling intensities for the stratigraphic intervals (although see refs 20 and 21), we fitted our model to two schemes (see Table 1 for details). In both schemes, the sampling intensity in the interval preceding the first to yield fossil primates was one-tenth of that in the first such interval. This option reflects the possibility that the first members of the clade

were unusually small²², perhaps had very low population sizes and/or had a limited geographical distribution and may thus be harder to find in the fossil record. It has been demonstrated that preservation rates for mammals were lower in the Cretaceous period than during the Cenozoic era²⁰.

Various parametrizations of the logistic diversification model were investigated, but the great diversity of holarctic primates during the Eocene suggests that 90% of modern diversity had already been reached by the Middle Eocene, 49 Myr ago. All analyses were run with the mean species duration set to 2.5 Myr (ref. 23), but the results change by only 1% if mean longevity is increased to 3 Myr or decreased to 2 Myr, so the method appears to be relatively insensitive to assigned species longevity.

To address how well our model describes the data, we used the fitted values of the X^2 statistic defined in equation (5) (see Methods). Scheme 1 gave a value of $X^2 = 116.8$, considerably larger than expected. Because of this lack of fit, this scheme is not discussed further. Scheme 2 gave a value of $X^2 = 28.5$, consistent with expectations; this scheme provides an adequate fit to the data. The estimated time of the LCA for Scheme 2 is 81.5 Myr ago with a 95% confidence interval of (72.0, 89.6) Myr ago. Estimates of the time of the LCA of primates based on 90% of modern diversity being reached by the base of the Miocene epoch are also considerably older than the oldest known fossil, lying even further in the past. The sampling probabilities up to the base of the Eocene have an average value of 0.057, with an upper 95% confidence limit of 0.074.

While our results agree broadly with a molecular estimate of the time of the strepsirrhine and haplorhine divergence⁵, they contradict widely accepted palaeontological estimates. Gingerich and Uhen²⁴ argued that, at a 95% confidence level, primates originated 55–63 Myr ago. However, their model gives the same results regardless of the number of modern primate species and regardless of species preservation rate. In fact, application of that model to a 6 Myr gap in the primate fossil record between the Early and the Late Oligocene epoch yields an extremely low probability for the existence of primates during that gap. Foote *et al.*²⁰ argue that, if molecular clock estimates of Cretaceous origins of living mammalian orders are correct, the preservation potential per lineage per million years (r) must be at least an order of magnitude smaller than they had estimated: between 0.25 and 0.37 (refs 20 and 25) for Cenozoic mammals and 0.03 for Cretaceous mammals. Our sampling intensities (α_i) divided by average species longevity are equivalent to r of ref. 20 and may be estimated from equation (6) (see Methods). The average values for preservation potential obtained using our approach—0.023 per lineage per Myr, and 0.003 per lineage per Myr for the last interval—are, indeed, an

order of magnitude smaller. Preservation rates proposed²⁰ for modern eutherian mammals are based either entirely (Cenozoic) or predominantly (Cretaceous) on North American faunas. North America is the best-sampled region in the world, and estimates based on that region will necessarily overestimate preservation rates of groups with a wider distribution.

The preservation rates of ref. 20 are likely to be overestimates for several reasons. Methods for assessing completeness based exclusively on the fossil record can only account for gaps within known lineages. Foote²⁵ demonstrated that the method used in ref. 20 will overestimate preservation potential where chronological gaps occur. The primate fossil record as a whole has several large gaps, most notably a general absence of fossils between the Early and Late Oligocene (at least 6 Myr) and the absence of a fossil record for the Malagasy lemurs, a diverse group with at least 46 living and subfossil species. Geographical gaps are equally substantial. Living primates are essentially confined to tropical and subtropical habitats²⁶. Primates populated substantial parts of the northern continents only when these areas supported subtropical habitats, during the Eocene and the Miocene. Yet 47% of known fossil primate species come from North America and Europe and, for the first half of palaeontologically documented primate evolution, sites yielding fossil primates are largely restricted to these two regions (Fig. 2).

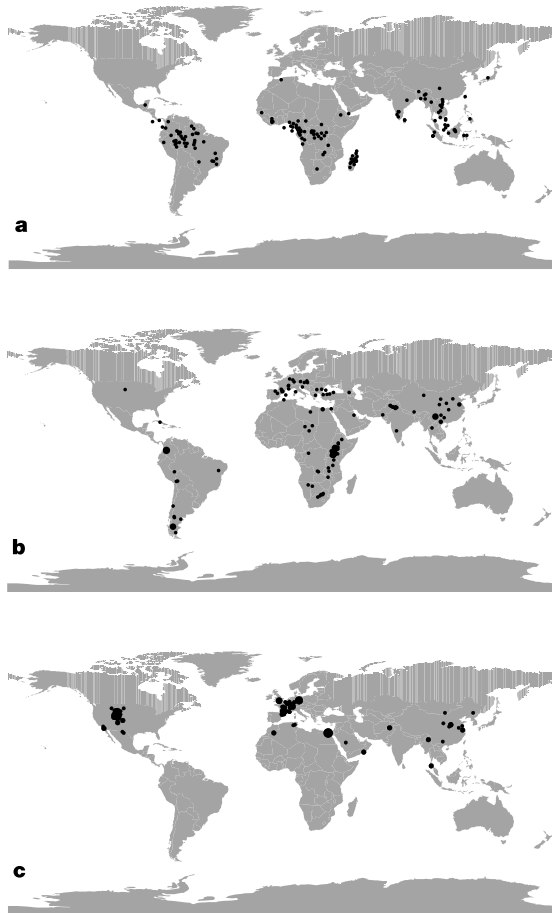


Figure 2 Mid-range of geographical distribution for individual modern and fossil primate species. **a**, Modern and sub-fossil primates (170 species obtained from Wolfheim's review of the distribution of modern primates²⁹). **b**, Fossil species for the Late Pleistocene to the Late Oligocene (167 species). **c**, Fossil species from the Early Oligocene to the Early Eocene (196 species). (The database of fossil primates was compiled from a large number of published sources. A full list of references can be found at <http://www.unizh.ch/anthro/Main/Who/Soligo/supinfo2.html>).

Direct reading of the known fossil record suggests that primates originated during the Palaeocene in the northern continents and subsequently migrated southwards. An alternative interpretation is that primates originated earlier in the poorly documented southern continents and expanded northwards when climatic conditions permitted. □

Methods

A model for speciation

The speciation model we use is the non-homogeneous Markov branching process^{13,14}. To model evolution after the LCA, we start with two species at time 0. Species have exponential lifetimes with mean $1/\lambda$, time being measured in millions of years. A species that goes extinct at time u is replaced by an average of $m(u)$ new species. Write Z_t for the number of species alive at time t , and define $B[s, t]$ to be the number of species that are born in the time interval $[s, t]$. The expected number of species extant at time t is given¹³ by

$$EZ_t = 2 \exp \left\{ \lambda \int_0^t (m(u) - 1) du \right\} \quad (1)$$

and

$$EB[s, t] = \lambda \int_s^t m(u) EZ_u du, \quad s < t \quad (2)$$

To model the mean diversification, we use the logistic function for which $EZ_t = 2/(\gamma + (1 - \gamma)e^{-\rho t})$.

A model for fossil finds

We divide time into k stratigraphic intervals, beginning from the present and proceeding into the past (see Table 1 and Fig. 1). The base of the first (youngest) stratigraphic interval is at T_1 Myr ago and the base of the k th is at T_k Myr ago. The earliest fossil is found in this interval. The founding species originate at time $T = T_1 + \tau$ Myr ago, and we define a $(k + 1)$ th stratigraphic interval that has its base at $T_{k+1} := T$ Myr ago and ends T_k Myr ago. No fossils have been found in this interval. We estimate the parameter τ , the temporal gap, using as data the number of different species found in the fossil record in the first, second, ..., k th intervals.

To do this, we model the number of species alive u Myr ago by the value Z_{T-u} of the Markov branching process. The number N_j of distinct species living in the j th stratigraphic interval having base T_j Myr ago is the sum of those that were extant at the beginning of the interval, Z_{T-T_j} plus those that originated in the interval $B[T - T_j, T - T_{j-1}]$. It follows from equations (1) and (2) that the expected number of distinct species that can be sampled in the j th stratigraphic interval is

$$EN_j = EZ_{T-T_{j-1}} + \lambda \int_{T-T_j}^{T-T_{j-1}} EZ_u du, \quad j = 1, \dots, k + 1 \quad (3)$$

We assume that, conditional on the number of distinct species N_j in the j th stratigraphic interval ($j = 1, 2, \dots, k + 1$), the number of species D_j actually found in the fossil record in this interval is a binomial random variable with parameters N_j and α_j , $j = 1, 2, \dots, k$. Furthermore, the D_j are assumed to be conditionally independent given the N_j . The parameter α_j gives the probability of sampling a species in the j th stratigraphic interval. Under this sampling model, the expected number of species found in the j th interval is

$$ED_j = \alpha_j EN_j, \quad j = 1, \dots, k + 1 \quad (4)$$

Statistical approach

Our method estimates τ by minimizing a weighted sum of squares of differences between the observed numbers of species found in the j th stratigraphic interval and the expected numbers given by equation (4). We assume that the number of species N_0 alive now is equal to the expected number under the diversification model; thus $N_0 = EZ_T$, which serves to determine one parameter of the diversification model.

Assuming that α_j is small, it can be shown that the variance of D_j is approximately $\alpha_j EN_j$. Our statistic therefore takes the form

$$X^2 = \sum_{j=1}^{k+1} (D_j - \alpha_j EN_j)^2 / \alpha_j EN_j \quad (5)$$

where $D_{k+1} = 0$ because no species are found from the earliest stratigraphic interval.

We also model the form of the sampling probabilities, α_j . One parsimonious choice is to set

$$\alpha_j = \alpha p_j, \quad j = 1, 2, \dots, k + 1 \quad (6)$$

where the p_j are known relative sampling intensities and the scale parameter α is to be estimated. Our estimates of τ and α are given by the values that minimize the quantity X^2 defined in equation (5). Estimates of the absolute sampling fractions can be obtained from equation (6) using the estimate of α . We can also estimate the average species longevity, but for simplicity we assume it is known here.

Bias and approximate confidence intervals

We can obtain approximate confidence intervals for τ and α using a parametric bootstrap approach; compare with Ch. 5 of ref. 27. We simulate b realizations of a birth–death

process Z_t starting from $Z_0 = 2$, using the values τ_0 and α_0 of τ and α estimated from the data as the parameters in each run. Paths in which either branch dies out before time T are excluded from the analysis. For each accepted run, i , we re-estimate τ and α , getting values $\tau_i, \alpha_i, i = 1, \dots, b$. Run i also produces a value X_i^2 of the statistic in equation (5). Assuming²⁷ that the values of $\tau_i - \tau_0$ mimic the distribution of $\tau_0 - \tau$, we obtain a bias-corrected estimate of τ of $2\tau_0 - \bar{\tau}$, where $\bar{\tau}$ is the sample average of the replicates $\tau_1, \dots, \tau_b$. An approximate $100(1 - 2v)\%$ equal-tailed confidence interval for τ is then given by $(2\tau_0 - \tau_{(b(1-v))}, 2\tau_0 - \tau_{(bv)})$, where $\tau_{(j)}$ is the j th largest of $\tau_1, \dots, \tau_b$. We used $b = 2,500$. A similar method is used to find an upper 95% confidence interval for α and the α_j . The empirical distribution of X^2 can be found from the values of $X_1^2, \dots, X_b^2$.

Received 18 September 2001; accepted 17 January 2002.

1. Janke, A., Feldmaier-Fuchs, G., Thomas, W. K., von Haeseler, A. & Pääbo, S. The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics* **137**, 243–256 (1994).
2. Easteal, S., Collett, C. & Betty, D. *The Mammalian Molecular Clock* (Landes, Austin, 1995).
3. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
4. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388**, 61–64 (1997).
5. Arnason, U., Gullberg, A. & Janke, A. Molecular timing of primate divergences as estimated by two nonprimate calibration points. *J. Mol. Evol.* **47**, 718–727 (1998).
6. Huelsenbeck, J. P., Larget, B. & Swofford, D. A compound Poisson process for relaxing the molecular clock. *Genetics* **154**, 1879–1892 (2000).
7. Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**, 610–614 (2001).
8. Murphy, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614–618 (2001).
9. Eizirik, E., Murphy, W. J. & O'Brien, S. J. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* **92**, 212–219 (2001).
10. Martin, R. D. Primate origins: plugging the gaps. *Nature* **363**, 223–234 (1993).
11. Martin, R. D. in *Major Topics in Primate and Human Evolution* (eds Wood, B. A., Martin, L. B. & Andrews, P.) 1–31 (Cambridge Univ. Press, Cambridge, 1986).
12. Marshall, C. R. in *The Adequacy of the Fossil Record* (eds Donovan, S. K. & Paul, C. R. C.) 23–53 (Wiley, London, 1998).
13. Harris, T. E. *The Theory of Branching Processes* (Springer, Berlin, 1963).
14. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* **344**, 305–311 (1994).
15. Raup, D. M., Gould, S. J., Schopf, T. M. & Simberloff, D. S. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* **81**, 525–542 (1973).
16. MacArthur, R. H. & Wilson, E. O. An equilibrium theory of insular zoogeography. *Evolution* **17**, 373–387 (1963).
17. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 1967).
18. Sepkoski, J. J. Jr A kinetic model of Phanerozoic taxonomic diversity: I. Analysis of marine orders. *Paleobiology* **4**, 223–225 (1978).
19. Alroy, J. in *Biodiversity Dynamics* (eds McKinney, M. L. & Drake, J. A.) 232–287 (Columbia Univ. Press, New York, 1998).
20. Foote, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. J. Jr Evolutionary and preservational constraints on origins of biological groups: divergence times of eutherian mammals. *Science* **283**, 1310–1314 (1999).
21. Foote, M. & Sepkoski, J. J. Absolute measures of the completeness of the fossil record. *Nature* **398**, 415–417 (1999).
22. Alroy, J. Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science* **280**, 731–733 (1998).
23. Alroy, J. Quantitative mammalian biochronology and biogeography of North America. PhD dissertation. Univ. Chicago (1994).
24. Gingerich, P. D. & Uhen, M. D. Time of origin of primates. *J. Hum. Evol.* **27**, 443–445 (1994).
25. Foote, M. Estimating taxonomic durations and preservation probability. *Paleobiology* **23**, 278–300 (1997).
26. Martin, R. D. *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Chapman Hall/Princeton Univ. Press, London/New Jersey, 1990).
27. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and their Applications* (Cambridge Univ. Press, Cambridge, 1997).
28. Wolfheim, J. H. *Primates of the World: Distribution Abundance and Conservation* (Univ. Washington Press, Seattle/London, 1983).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank U. Arnason, A. Müller and U. Thalmann for helpful comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.T. (e-mail: stavare@usc.edu).

Fisheries productivity in the northeastern Pacific Ocean over the past 2,200 years

Bruce P. Finney^{*†}, Irene Gregory-Eaves^{†‡}, Marianne S. V. Douglas[§] & John P. Smol[‡]

^{*} Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, Alaska 99775-7220, USA

[‡] Paleoecological Environmental Assessment and Research Laboratory (PEARL), Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

[§] Paleoenvironmental Assessment Laboratory (PAL), Department of Geology, University of Toronto, 22 Russell St., Toronto, Ontario M5S 3B1, Canada

[†] These authors contributed equally to this work

Historical catch records suggest that climatic variability has had basin-wide effects on the northern Pacific and its fish populations, such as salmon, sardines and anchovies^{1–7}. However, these records are too short to define the nature and frequency of patterns. We reconstructed ~2,200-year records of sockeye salmon abundance from sediment cores obtained from salmon nursery lakes on Kodiak island, Alaska. Large shifts in abundance, which far exceed the decadal-scale variability recorded during the past 300 years^{1–8}, occurred over the past two millennia. A marked, multi-centennial decline in Alaskan sockeye salmon was apparent from ~100 BC to AD 800, but salmon were consistently more abundant from AD 1200 to 1900. Over the past two millennia, the abundances of Pacific sardine and Northern anchovy off the California coast, and of Alaskan salmon, show several synchronous patterns of variability. But sardines and anchovies vary out of phase with Alaskan salmon over low frequency, which differs from the pattern detected in historical records^{5,6}. The coherent patterns observed across large regions demonstrate the strong role of climatic forcing in regulating northeastern Pacific fish stocks.

Salmon are important ecological, economical and cultural resources in the northern Pacific region, and their response to future climatic change is very uncertain⁹. Long-term relationships between sockeye salmon populations and climatic change can be evaluated by analysing sediment cores from their nursery lakes⁸. After one to four years of feeding in the northern Pacific, sockeye salmon (*Oncorhynchus nerka*) return to their natal lake/stream system to spawn and die¹⁰. The nutrients derived from spawned carcasses can be significant relative to other sources, and may be reconstructed from palaeolimnological records of $\delta^{15}\text{N}$ and algal bioindicators, such as diatoms⁸. Periods of greater input of salmon-derived nutrients (SDN), and hence greater sockeye salmon abundance, corresponded to higher sedimentary $\delta^{15}\text{N}$ and more eutrophic diatom taxa. Here we show reconstructions of salmon abundance over the past two millennia from lakes on Kodiak island, Alaska (Fig. 1), where our proxies have been successfully calibrated with monitoring data⁸. This region is one of the most important salmon-producing areas of the northern Pacific, and historical records suggest that its salmon abundances are representative of population trends in Alaska⁴.

In Karluk lake (57°25' N, 154°05' W), our sedimentary indicators of SDN— $\delta^{15}\text{N}$ and diatoms—show striking changes and a strong degree of coherence ($r^2 = 0.79$, $n = 100$, $P < 0.001$) over the past ~2,200 years (Fig. 2). In the oldest sediments, around 200 BC, we infer from the high $\delta^{15}\text{N}$ values and the strong presence of mesotrophic to eutrophic diatom taxa (for example, *Stephanodiscus minutulus*/parvus) that the return of sockeye salmon to Karluk lake was high, and similar to levels (~ 3 million yr^{-1}) present when

letters to nature

commercial fishing began at the Karluk estuary in AD 1882 (ref. 8). A striking feature of the record is the dramatic and sustained change to lower SDN loading beginning at ~ 100 BC, as inferred from a shift in the diatom community to an oligotrophic assemblage (that is, a change in dominants to *Cyclotella comensis* and *C. ocellata*) and the strong decrease in $\delta^{15}\text{N}$. Beginning at about AD 250, there is a general

increase in $\delta^{15}\text{N}$ and an increase in oligo- to mesotrophic diatom taxa (for example, *Cyclotella pseudostelligera* and *Fragilaria crotonensis*), but peak abundance in SDN is not reached until $\sim$ AD 1200. Overall high levels of SDN are inferred for the period AD 1200–1900, with decadal-scale shifts as the main source of variability during this portion of the record. Declines in SDN since the 1900s can be

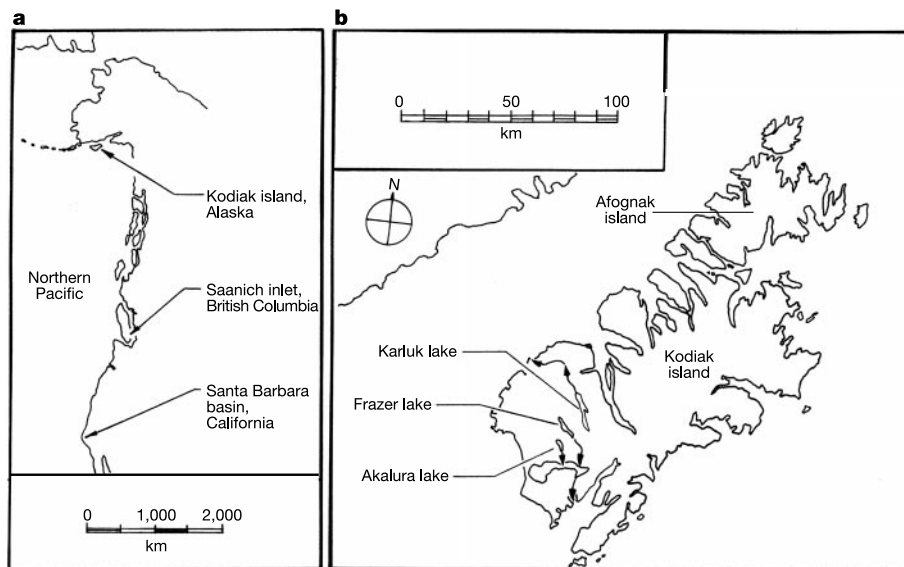


Figure 1 a, Map of the northeastern Pacific coast from Alaska to California, highlighting Kodiak island (this study), Saanich inlet²⁷ and the Santa Barbara basin^{18,19,25}. **b**, Map of Kodiak Island showing Karluk, Akalura and Frazer lakes. That these lakes drain to different

parts of the island reinforces the concept that the $\delta^{15}\text{N}$ pattern reproduced in Karluk and Akalura lakes is not caused by tectonic processes.

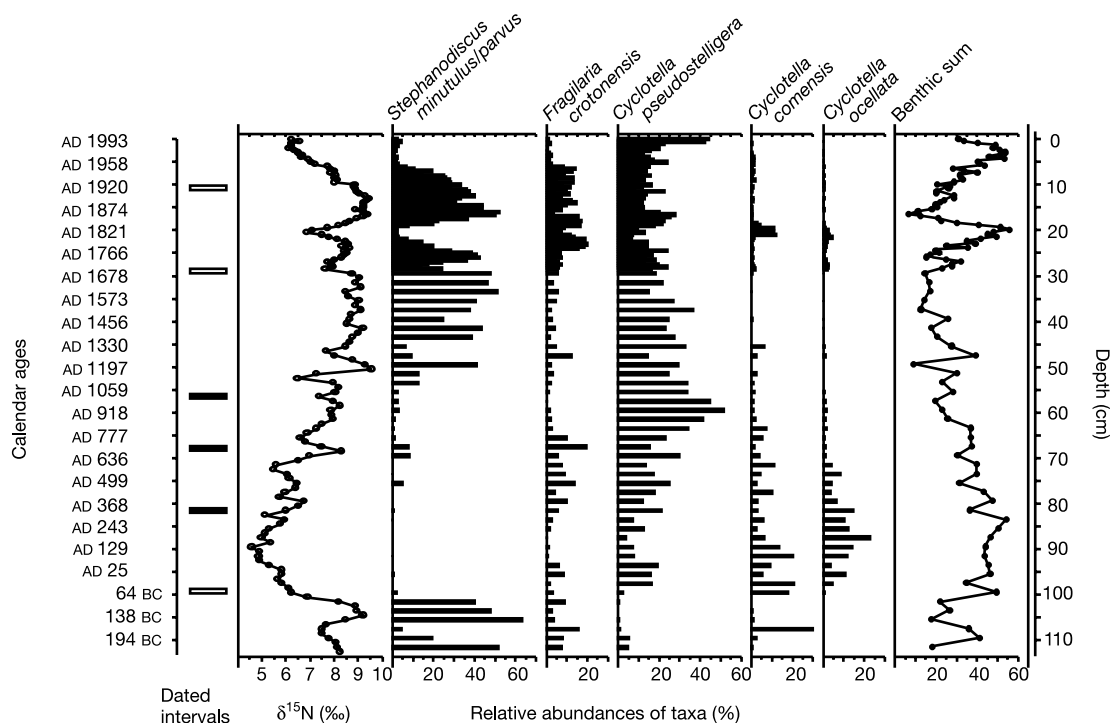


Figure 2 Palaeolimnological evidence of dramatic changes in sockeye salmon abundances from Karluk lake over the past $\sim 2,200$ years. As salmon transport significant quantities of nutrients enriched in $\delta^{15}\text{N}$ relative to terrestrial sources, high $\delta^{15}\text{N}$ values and the strong presence of *Stephanodiscus minutulus/parvus* (a mesotrophic to eutrophic indicator) reflect phases of large returns of sockeye salmon to Karluk lake. In contrast, periods of lower salmon-derived nutrients are marked by low $\delta^{15}\text{N}$ values and oligotrophic

(that is, *Cyclotella comensis* and *C. ocellata*) and benthic taxa⁸. Intermediate salmon-derived nutrient conditions are indicated by *C. pseudostelligera* and *Fragilaria crotonensis* (which have been shown to be responsive to moderate lake fertilization in Pacific coast lakes)^{8,29}. The age model for this record was derived from dated tephras (white bars) and calibrated ^{14}C dates (black bars).

attributed to commercial fishing and climatic changes⁸.

Direct impacts of climatic changes, and local landscape changes resulting from volcanism, tectonic activity and fires, can be ruled out as the primary sources of variability in the Karluk lake record. The volcanic ash layers are relatively thin (<1 cm) and significantly lag $\delta^{15}\text{N}$ and diatom changes. Tectonic processes can be ruled out on the basis of palaeoseismicity studies that suggest this region is subsiding, not uplifting, which would be a mechanism more likely to restrict access to the lakes¹¹. The lack of a strong response in our nearby reference lake, Frazer lake (57° 15' N, 154° 08' W), over the past ~2,200 years suggests that the direct effects of climatic change (independent of sockeye salmon population fluctuations) were not the primary factors influencing Karluk lake. Frazer lake is very similar to Karluk lake, but has a steep waterfall at its outlet, which naturally prohibited salmon migration. It was not until the 1950s, when eggs were planted and a fish ladder was constructed, that a self-sustaining sockeye salmon run was established in Frazer lake¹². Therefore any changes evident in the Frazer lake palaeolimnological record before the 1950s would reflect only climatic and landscape shifts. However, the $\delta^{15}\text{N}$ signature of Frazer lake is low and shows little variation over the past ~2,200 years, as would be expected for a reference lake with no influence of SDN (Fig. 3). Diatom analysis of these sediments also reveals only subtle variation and reflects constant oligotrophic conditions, consistent with no introduction of SDN. The greatest changes in the Frazer lake $\delta^{15}\text{N}$ and diatom records are associated with the introduction of sockeye salmon in recent decades⁸.

The sustained multi-centennial regimes of anomalous salmon abundance, unprecedented during the historical period or during our short-term 300-year records⁸, appear to be regional phenom-

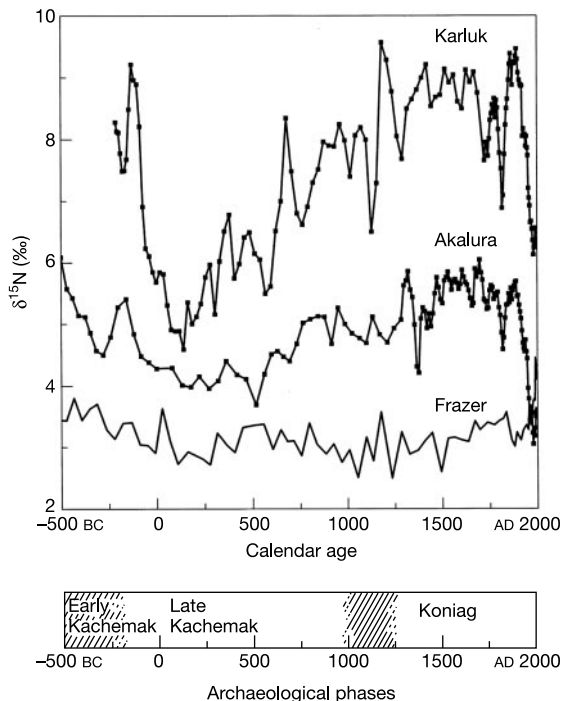


Figure 3 A regional comparison of sedimentary $\delta^{15}\text{N}$ profiles and the archaeological phases¹⁰ from Kodiak island, Alaska. Both Karluk and Akalura lakes are natural sockeye salmon nursery lakes. Frazer lake, our reference lake, was without sockeye salmon until the 1950s (when it was stocked and a fish ladder was constructed) because it has an impassable waterfall at its outlet. The importance of changes in sockeye salmon abundance to the indigenous peoples of Kodiak is reflected in archaeological deposits¹⁴ and the delineation of cultural phases¹³. A shift towards greater abundance of fishing tools¹⁴, greater housing densities and larger multi-room houses is evident during the Koniag phase^{14,30}.

ena, because a similar pattern to that seen in Karluk lake is observed in a second sockeye nursery lake, Akalura lake (57° 11' N, 154° 12' W; Fig. 3). The reconstructed patterns of salmon abundance also have strong similarities to changes in cultural characteristics and population estimates determined from archaeological data from Kodiak island. The abrupt reduction in SDN at ~100 BC appears to coincide with a major cultural change associated with the transition from the early to late Kachemak period¹³ (Fig. 3). Subsequently, the increase in SDN from AD 800 to AD 1200 matches the change from the late Kachemak to the transitional Koniag and Koniag periods, when population numbers rose and a strong shift towards high utilization of salmon fishing gear is evident¹⁴. These archaeological data suggest that natural variability in salmon abundance influenced human culture, which differs from what has been observed in the twentieth century⁸, and in other fisheries¹⁵, where anthropogenic fishing has been an important determinant of coastal marine fish production.

Analysis of salmon catch records since the 1920s (refs 1–5, 7) strongly suggest that physical changes in the north Pacific Ocean are an important control on salmon production. Monitoring of primary and secondary production in the northeastern Pacific^{16,17}

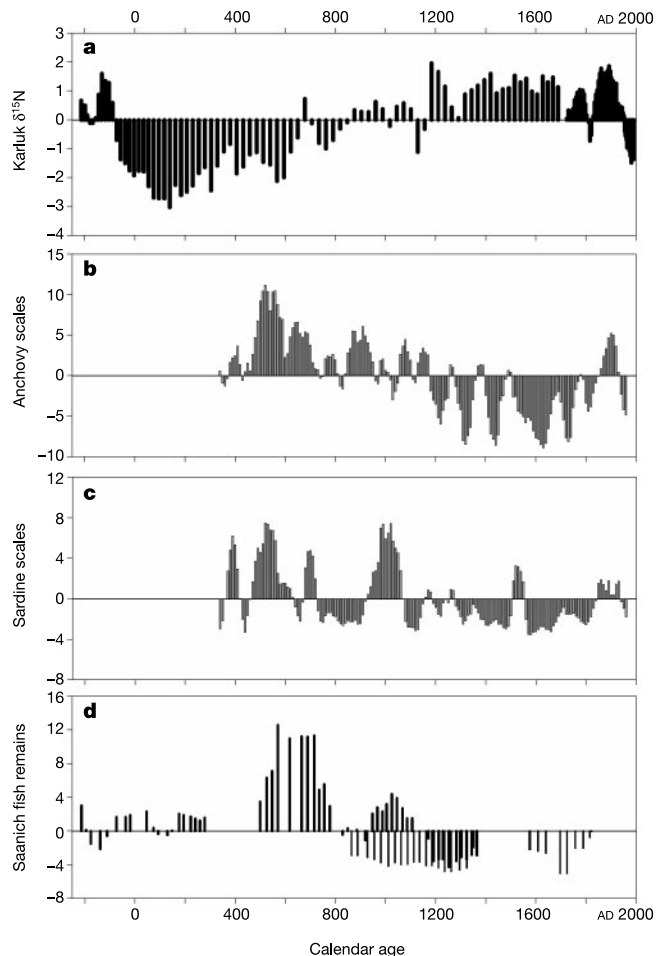


Figure 4 Reconstructions of fish abundances for the northeastern Pacific Ocean over the past ~2,200 years. Each series is plotted as the difference from the series mean calculated over the time period presented. **a**, The Karluk lake $\delta^{15}\text{N}$ profile (‰) as a proxy for sockeye salmon abundances in Alaska. **b**, **c**, Northern anchovy and Pacific sardine scales (no. of scales per 1,000 cm² per year) from the Santa Barbara basin, California²⁵. A 50-year running average was applied to highlight long-term trends. **d**, Fish remains (including Pacific herring, Pacific hake and cartilaginous fish) (no. of fish remains per 100 cm³) recovered from Saanich inlet, British Columbia²⁷; data from two overlapping cores are presented.

indicates that climatic forcing has a direct impact on lower trophic levels, which subsequently affects salmon production⁴. Historically, significant changes in states of northern Pacific fish populations and climate have occurred over interdecadal-scale periods^{1–7}. However, our long-term data reveal the existence of longer multi-centennial regimes. The two noticeable multi-century shifts in inferred sockeye abundance at ~100 BC and ~AD 800–1200 correspond to periods of major change in ocean–atmosphere circulation in the northeastern Pacific. The dramatic decrease in Alaskan salmon abundance at ~100 BC is contemporaneous with warming of marine waters in the Santa Barbara basin, California^{18,19}. Increased abundances of Alaskan salmon after ~AD 1200 correspond to a period of glacial advances in southern Alaska²⁰ and the Canadian Rockies²¹, whereas high-resolution palaeoclimatological data from western North America show a dramatic shift at ~AD 800–1000 to drier conditions^{22,23}.

Twentieth-century monitoring of primary and secondary production and fish catch data also suggest that marine productivity along the west coast of North America is roughly divided by the Subarctic Current, which flows eastward to meet the coast near the Queen Charlotte Islands. Oscillating modes of productivity between the northern versus southern waters have been observed historically^{3,24}. Our reconstruction of Alaskan salmon abundance is out of phase with the low-frequency long-term trends in both Pacific sardines (*Sardinops sagax*) and Northern anchovies (*Engraulis mordax*) over the past 1,700 years from the Santa Barbara basin²⁵ (Fig. 4). From ~AD 300 to AD 1200, sardines and anchovies were more abundant when Alaskan sockeye stocks were much weaker than average, whereas a reversal of this trend is apparent for the last ~800 years. The relatively high abundances of both sardines and anchovies²⁵, as well as proxies of upwelling and productivity from the Santa Barbara basin¹⁹ and Gulf of California²⁶, indicate that the coastal zone off California was more productive from ~AD 300 to AD 1200. Patterns inferred from fish bones deposited in southern coastal British Columbian sediments indicate that fish productivity was also higher from ~0 to AD 1000 (ref. 27, Fig. 4). Thus the pattern of opposite trends in Alaskan salmon and the Pacific Northwest to California fish productivity was sustained for long periods, and was of greater amplitude than regimes in the twentieth century. The strong coherence of changes in regional fish populations and palaeoclimatic trends supports the hypothesis of linkages between ocean–atmosphere circulation and northern Pacific ecosystems, and suggests major changes in modes at ~BC 100 and AD 1000–1200.

The long-term trend reported here is distinct from the pattern seen in historical records, where sardines covary with salmon, but are out of phase with anchovies^{5,6}. An additional pattern of variability not apparent in catch data, but revealed through our high-resolution palaeoecological analysis, is one in which sardines, anchovies and Alaskan salmon undergo synchronous in-phase fluctuations, such as those observed from the early 1800s through to the mid-1900s (Fig. 4). These relationships, revealed on different timescales, suggest that multiple modes of variability in ocean–atmosphere circulation and ecosystem dynamics operate in the northern Pacific. The unique low- and high-frequency relationships that we observe highlight the value of long records in identifying patterns in fish stock dynamics.

Contemporary monitoring and catch records have provided only a limited understanding of fish population dynamics and their responses to climatic change. However, palaeolimnological and palaeoceanographic investigations can greatly expand our knowledge, and provide an essential perspective on the magnitude and duration of fish stock variability. Our ~2,200-year reconstructions of Alaskan sockeye salmon abundances demonstrate that an unprecedented shift to a very low productivity regime, lasting centuries, can occur even without the influence of fisheries and other anthropogenic impacts. Synchronous variability between the few other

palaeoceanographic records and our data strongly suggests that climate-related factors may greatly influence fish abundances over large geographic areas. A more thorough understanding of the linkages between climatic change and ocean ecosystems is critical for future sustainable management of northern Pacific fisheries, as fish stocks are now faced with many additional stresses including commercial fishing, habitat degradation and global warming. □

Methods

Undisturbed surface sediment cores were obtained from the lakes using a hand-operated gravity corer with a messenger-activated valve, specifically developed for collecting high-resolution surface cores. Age control for the past ~300 years is based on ²¹⁰Pb and ¹³⁷Cs dating and identification of volcanic ash layers⁸. Longer cores were obtained with a percussion corer, and the records were spliced together at a volcanic ash layer dated to ~AD 1710. Age control before this ash is based on accelerator mass spectrometry (AMS) radiocarbon dating of terrestrial macrofossils, and correlation of volcanic ash layers. A predominant ash found in all cores has been dated to 20 BC in the Frazer lake core. For Karluk lake, three additional AMS dates were obtained above this ash, and an age model was constructed using a third-order polynomial fit ($r^2 = 0.99$) with these four dated intervals and the two ash layers found in the surface cores (Katmai AD 1912, and the ~AD 1710 ash). The Akalura and Frazer lakes age models were constructed from regression equations ($r^2 > 0.95$) using the three ash dates. The character of the sediment cores is uniform throughout (except for the ashes), suggesting no hiatuses or anomalies (such as turbidites) in the depositional process.

Cores were sampled continuously at 0.5-cm intervals. Nitrogen isotopes were measured on dried, homogenized, bulk samples using a Finnigan Delta Plus mass spectrometer. Isotopic analyses are reported in standard δ notation relative to atmospheric N₂. Analytical precision is better than $\pm 0.2\%$. Isotope analysis of replicated cores collected from Karluk lake shows excellent reproducibility of the coring, isotope and dating techniques. Diatom slides were prepared following standard methods²⁸. At least 400 valves were identified and enumerated along central transects of the coverslip for each sample, using a Leitz DMRB microscope equipped with differential interference contrast (numerical aperture, 1.30; magnification, 1000 $\times$).

Received 16 November 2001; accepted 14 January 2002.

1. Beamish, R. J. & Bouillon, D. R. Pacific salmon production trends in relation to climate. *Can. J. Fish. Aquat. Sci.* **50**, 1002–1016 (1993).
2. Francis, R. C. & Hare, S. R. Decadal-scale regime shifts in the large marine ecosystem of the north-east Pacific: a case for historical science. *Fish. Oceanogr.* **3**, 279–291 (1994).
3. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
4. Francis, R. C., Hare, S. R., Hollowed, A. B. & Wooster, W. S. Effects of interdecadal climate variability on the oceanic ecosystems of the NE Pacific. *Fish. Oceanogr.* **7**, 1–21 (1998).
5. Beamish, R. J. *et al.* The regime concept and natural trends in the production of Pacific salmon. *Can. J. Fish. Aquat. Sci.* **57**, 516–526 (1999).
6. Luch-Belda, D. in *Large Marine Ecosystems of the North Pacific Rim* (eds Sherman, K. & Tang, Q.) 42–47 (Blackwell Science, Malden, Massachusetts, 1999).
7. Hare, S. R. & Mantua, N. J. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Prog. Oceanogr.* **47**, 103–145 (2000).
8. Finney, B. P., Gregory-Eaves, I., Sweetman, J., Douglass, M. S. V. & Smol, J. P. Impacts of climatic change and fishing on Pacific salmon abundances over the past 300 years. *Science* **290**, 795–799 (2000).
9. Welch, D. W., Ishida, Y. & Nagasawa, K. Thermal limits and ocean migrations of sockeye salmon (*Oncorhynchus nerka*): long-term consequences of global warming. *Can. J. Fish. Aquat. Sci.* **55**, 937–948 (1998).
10. Burgner, R. L. in *Pacific Salmon Life Histories* (eds Groot, C. & Margolis, L.) 3–117 (UBC Press, Vancouver, 1991).
11. Gilpin, L. M. *Holocene Paleoseismicity and Coastal Tectonics of the Kodiak Islands, Alaska*. PhD thesis, Univ. California Santa Cruz (1995).
12. Burger, C. V., Scribner, K. T., Swanton, C. O. & Campton, D. E. Genetic contribution of three introduced life history forms of sockeye salmon to colonization of Frazer Lake, Alaska. *Can. J. Fish. Aquat. Sci.* **57**, 2096–2111 (2000).
13. Mills, R. O. Radiocarbon calibration of archaeological dates from the central Gulf of Alaska. *Arctic Anthropol.* **31**, 126–149 (1994).
14. Kneet, R. *The Late Prehistory of the Alutiiq People: Culture Change on the Kodiak Archipelago from 1200–1750 AD*. PhD thesis, Bryn Mawr College, Pennsylvania (1995).
15. Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–638 (2001).
16. Venrick, E. L., McGowan, J. A., Cayan, D. R. & Hayward, T. L. Climate and chlorophyll *a*: long-term trends in the central north Pacific Ocean. *Science* **238**, 70–72 (1987).
17. Brodeur, R. D. & Ware, D. M. Long-term variability in zooplankton biomass in the subarctic Pacific Ocean. *Fish. Oceanogr.* **1**, 32–38 (1992).
18. Pisias, N. G. Paleoclimatology of the Santa Barbara Basin during the last 8000 years. *Quat. Res.* **10**, 366–384 (1978).
19. Kennett, J. P. & Ingram, B. L. A 20,000-year record of ocean circulation and climate change from the Santa Barbara basin. *Nature* **277**, 510–513 (1995).
20. Calkin, P. E., Wiles, G. C. & Barclay, D. J. Holocene coastal glaciation of Alaska. *Quat. Sci. Rev.* **20**, 449–461 (2001).
21. Luckman, B. H., Briffa, K. R., Jones, P. D. & Schweingruber, F. H. Tree-ring based reconstructions of summer temperatures at the Columbia Icefield, Alberta, Canada, AD 1073–1983. *Holocene* **7**, 375–389 (1997).
22. Bennett, J. *Paleoclimatic Investigations from a Closed-basin Lake in Southern Interior British Columbia*. MSc thesis, Queen's Univ. Ontario (1999).

23. Pienitz, R., Smol, J. P., Last, W. M., Leavitt, P. R. & Cumming, B. F. Multi-proxy Holocene palaeoclimatic record from a saline lake in the Canadian subarctic. *Holocene* **10**, 673–686 (2000).
24. Gargett, A. E. The optimal stability 'window': a mechanism underlying decadal fluctuations in North Pacific salmon stocks? *Fish Oceanogr.* **6**, 109–177 (1997).
25. Baumgartner, T. R., Soutar, A. & Ferreira-Bartrina, V. Reconstruction of the history of Pacific sardine and northern anchovy populations over the last two millennia from sediments of the Santa Barbara Basin, California. *CalCOFI Rep.* **33**, 24–40 (1992).
26. Juillet-Leclerc, A. & Schrader, H. Variations of upwelling intensity recorded in varved sediment from the Gulf of California during the past 3000 years. *Nature* **329**, 146–149 (1987).
27. Tunnicliffe, V., O'Connell, J. M. & McQuoid, M. R. A Holocene record of marine fish remains from the Northeastern Pacific. *Mar. Geol.* **174**, 197–210 (2001).
28. Wilson, S. E., Cumming, B. & Smol, J. P. Assessing the reliability of salinity inference models from diatom assemblages: an examination of a 219-lake data set from western North America. *Can. J. Fish. Aquat. Sci.* **53**, 1580–1592 (1996).
29. Ennis, G. L., Northcote, T. G. & Stockner, J. G. Recent trophic changes in Kootenay Lake, British Columbia, as recorded by fossil diatoms. *Can. J. Bot.* **61**, 1983–1992 (1983).
30. Clark, D. W. Kodiak Island: the later cultures. *Arctic Anthropol.* **35**, 172–186 (1988).

Acknowledgements

We thank D. Barto, T. Chatto, R. Hander, S. Honnold, S. Schrof, N. Sagalkin and J. Sweetman for assistance in the field, and T. Brown, N. Haubenstein, A. Hirons, T. Howe, A. Krumhardt and M. Luoma for laboratory assistance. We are also grateful to J. Glew for drafting our map. B. Cumming, J. Dower, T. Johnston, D. Mann, and members of PEARL provided comments on the manuscript. This work was funded in the USA by the US Global Ocean Ecosystems Dynamics (GLOBEC) programme jointly funded by the National Science Foundation and the National Oceanic and Atmospheric Administration (NOAA), the Alaska Sea Grant College programme and the NOAA-Auke Bay Lab Ocean Carrying Capacity programme. In Canada, this research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ontario Graduate Scholarship (OGS) programme and the Northern Studies Training Program (NSTP).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.P.F. (e-mail: finney@ims.uaf.edu).

Sexual conflict reduces offspring fitness in zebra finches

Nick J. Royle^{*†}, Ian R. Hartley^{*} & Geoff A. Parker[‡]

^{*} School of Biological Sciences, IENS, University of Lancaster, Lancaster LA1 4YQ, UK

[‡] Population & Evolutionary Biology Research Group, School of Biological Sciences, University of Liverpool, Liverpool L69 3GS, UK

Parental care is often costly¹; hence, in sexually reproducing species where both male and female parents rear their offspring (biparental care), sexual conflict over parental investment can arise². Such conflict occurs because each care-giver would benefit from withholding parental investment for use with another partner, leading to a reduction in the amount of care given by one parent at the expense of the other^{3–5}. Here we report experiments to explore the prediction from theory that parents rearing offspring alone may provide greater parental investment than when rearing offspring together with a partner^{3,5}. We found that when the number of offspring per parent, and hence the potential workload, were kept constant, offspring received a greater per capita parental investment from single females than from both parents working together, and that males reared by single mothers were more sexually attractive as adults than their biparentally reared siblings. This difference between single- and two-parent families is due to a reduction in care provided by

females when they care together with a male, rather than laziness by males or differences in the begging behaviour of chicks, supporting the claim that sexual conflict in biparental care can reduce the quality of offspring raised^{3,5}.

The importance of conflicts in evolutionary processes has been increasingly recognized in recent years^{6–9}. In particular, sexual conflict over mating (pre-zygotic conflict) or parental investment (post-zygotic conflict) may be a powerful force shaping the potential for sexual selection^{2,10}, speciation¹¹ and the determination of life-history characteristics¹². Although pre-zygotic sexual conflicts such as sperm competition are well characterized^{13–17}, there are very few demonstrations of the effects of post-zygotic sexual conflict on offspring fitness¹⁸. We used zebra finches (*Taeniopygia guttata*) to explore theoretical predictions from models of sexual conflict^{3,5}, which suggest that in some circumstances, single parents should provide greater parental investment per chick and at greater cost to themselves, compared with rearing chicks with a partner.

Fourteen pairs of zebra finches were allocated equally to one of two groups once clutches had hatched. Females in each group raised one brood of four chicks with the male, and one brood of two chicks alone, but the order in which this was done was different in the two groups. In group 1, the male was removed by replacing the cage partition and placing him in the half away from the nest and female (see Methods). Brood size was maintained at, or reduced to, two when the chicks were 4–5 days old, by which time chicks are able to self-thermoregulate. The female then reared these chicks alone until they reached independence at 35 days, when the young were removed to a separate cage. This was the 'uniparental care' regime. The male and female were then re-united by the removal of the cage partition, and were allowed to start a second clutch. Both parents then reared the brood, which was adjusted as necessary so that the pair raised twice as many offspring (four) as the female had reared on her own (two). This was the 'biparental care' regime. In group 2, the order was reversed so that the biparental care regime preceded the uniparental care regime.

Consequently, in both of the groups each female experienced both treatments (uniparental and biparental care) consecutively, so that any effects of variation in parental investment on offspring fitness were not confounded by genetic effects. However, unlike previous male-removal experiments^{19–23}, which have generally shown that a single mother is unable to provide enough food for the full brood, we also simultaneously reduced brood size so that the number of chicks per parent remained constant. As the shape of the curve relating chick fitness to parental investment is likely to be set by the number of chicks per parent, any differences in parental

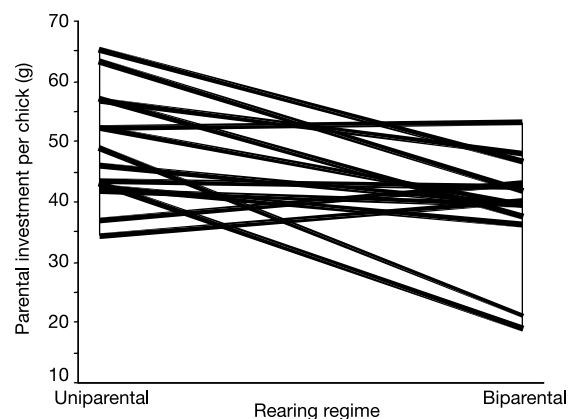


Figure 1 Parental investment for uniparental and biparental rearing regimes over a 15 day period after manipulation. Lines represent individual females. Our measure of parental investment, the amount of food consumed per chick, was greater under uniparental care (repeated measures general linear model, $F_{1,12} = 9.93$, $P = 0.008$).

[†] Present address: Division of Environmental Biology, Graham Kerr Building, Glasgow University, Glasgow G12 8QQ, UK.

Box 1

Theoretical framework

Historical background

Smith and Fretwell²⁷ analysed optimal investment by a single parent with a fixed effort to expend on progeny. This approach was extended by ref. 3 assuming the fixed effort to be a fixed lifetime's parental investment, comparing evolutionarily stable strategy (ESS) investments of uniparental and biparental care species. Additional studies^{25,28,29} examined ESS parental investment for species with biparental care, costing current expenditure in terms of lost future reproductive success. All of these approaches represent 'sealed bid' games: a parent is programmed to play a parental investment strategy independent of its mate's level of parental investment. Recently, ref. 30 modelled 'negotiation' games in which each parent responds in an ESS way to the current parental investment level of its mate, giving different ESS parental investment response levels from the 'sealed bid' approach.

Uniparental versus biparental care

Few theoretical analyses ask how offspring should fare when cared for by one parent rather than two. There are two obvious biological applications for this question, depending on whether parents are always either uniparental or always biparental (for example, comparison across species), or whether they face both situations during life, depending on circumstances (for example, uniparental care arising owing to mortality of a mate in a predominantly biparental species). One study³ (see also ref. 24) concluded that in two otherwise equivalent species, one biparental and the other uniparental, where the parents are equal and where brood size does not affect the relation between an offspring's success and the parental investment it receives, offspring should fare worse in the biparental species because of sexual conflict. However, this solution, although an ESS, is not continuously stable⁵, and this prediction cannot hold unless some modification can generate continuous stability. McNamara *et al.*⁵ compared uniparental with biparental care, for a given brood size, in a single species. Most models (for example, ref. 29) result in more total parental investment being supplied under biparental care, although for a given parent, parental investment under uniparental care should be greater. This fits with many of the experimental manipulations in which one parent is removed²⁰. However, certain cases (where parents are unequal, or there is negotiation) can result in less total parental investment in biparental situations.

Present experiment

The conclusions of ref. 5 relate to the case where a brood faces uniparental or biparental care, as could happen in nature. However, the instantaneous workload on a parent is greater if it must care for all the chicks on its own, rather than biparentally. Our experiments examine what happens when one parent rears two chicks singly, or shares equally with a partner the rearing of four chicks. The potential workload for a given parent is therefore equal in each situation. Note that in zebra finches, two and four are both naturally occurring brood sizes, and although parents are typically biparental, natural mortality is high, so that offspring are sometimes reared uniparentally. Conditions in the experiments can, therefore, be met in nature. Assuming that the behaviour we have observed is evolutionarily stable, this leaves two explanations for higher chick benefits under uniparental versus biparental care: parents are unequal and costs of parental investment are less for females, or conflict involves negotiation between parents, and the responsiveness of a given parent to a deficit by its mate is high. Although exact equality is unlikely, parental zebra finches are remarkably similar in their parental investment patterns. Furthermore, simple calculations (J. McNamara, personal communication) along 'sealed bid' lines suggest that if the costs of parental investment are less for the female, so that the two uniparental chicks fare better than the four biparental chicks, the female puts in more total effort with the four young than she does with the two (the male's effort is less than the female's and that is why the young fare worse with two parents). We found that females actually invest less in four chicks than in two, so the only explanation that seems compatible with our observations is that it represents the outcome of conflict between parents through negotiation.

investment between uniparental and biparental regimes will be due to the effects of sexual conflict, rather than just being a consequence of a reduction in clutch or brood size (Box 1 reviews the current theoretical perspective).

Parental investment per chick (see Methods) on the day before experimental manipulation was similar in uniparental and biparental regimes (paired *t*-test, $t_{13} = 0.62$, $P = 0.55$). However, after manipulation, for the period between the start of the experiment and fledging, chicks raised under a uniparental regime received 25% more parental investment than their siblings raised by two parents (Fig. 1). The factor 'group' and its interaction with rearing regime were both nonsignificant ($P > 0.50$). As any difference in the predicted direction could be just a consequence of males working less hard whilst females maintain the same effort between treatments, it was necessary to assess relative male and female provisioning effort. Data from videos showed that males, rather than providing less parental investment, actually provide similar or greater amounts than females rearing chicks under a biparental regime. The number of feeds per hour was not different between male and female parents, or for females between uniparental versus biparental regimes ($P > 0.50$, paired *t*-tests), but the mean feed duration (load size) was greater in males than in females ($t_{11} = 2.38$, $P = 0.037$). In addition, females raising chicks under a uniparental regime also provided feeds of a greater size than when raising chicks under a biparental regime ($t_{11} = 2.34$, $P = 0.039$). This provided confirmation that females increased their workload when rearing broods alone.

As the definition in ref. 2 makes clear, the currency of parental investment is its cost in terms of the parent's ability to invest in other offspring. In the current experiment, the costs of reproduction were felt by parents of both sexes, although these were expressed in different ways. Males lost over 6% of their mass between pairing and brood independence when helping to rear broods under biparental care (paired *t*-test, $t_{13} = -2.97$, $P = 0.012$). In contrast, when separated from the female while she reared a brood alone, males increased their mass by 6% over the same period ($t_{13} = 2.25$, $P = 0.043$). Females did not change mass after biparental or uniparental regimes ($P = 0.17$ and 0.85 , respectively; paired *t*-tests). Rather, investment in clutches (clutch mass) after a biparental care regime was over 20% greater than that after a uniparental regime (Fig. 2). This comprised an increase in both egg mass ($t_{13} = 2.50$, $P = 0.027$) and clutch size ($t_{13} = 3.31$, $P = 0.006$), and indicated greater parental investment by females raising broods alone.

The greater parental investment of single mothers had important consequences for offspring fitness. Chicks reared under biparental

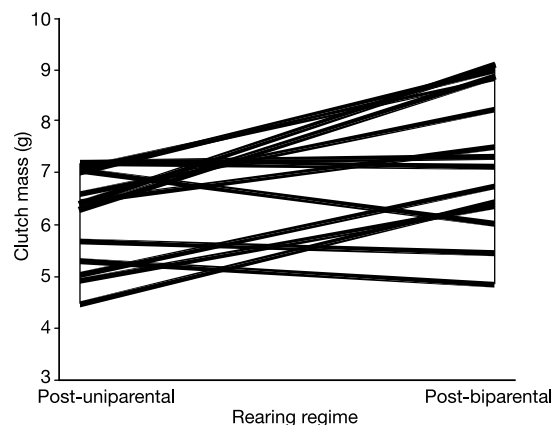


Figure 2 The mean mass of second clutches laid after the uniparental rearing regime (post-uniparental) is less than that of the biparental rearing regime (post-biparental; paired *t*-test, $t_{13} = 3.39$, $P = 0.005$). Lines represent individual females.

and uniparental care regimes were of a similar size (head–bill, tarsus and wing length) and mass at all ages of measurement after fledging (all $P > 0.20$; paired t -test comparing brood means), but as adults, male chicks differed in their attractiveness to the opposite sex. In a mate choice experiment, the males raised under the uniparental care regime were more attractive to females than were their brothers raised under the biparental care regime (Fig. 3a). Additionally, of the 12 females used in the test, only three did not show an overall preference for males raised under the uniparental regime (Fig. 3b), and one of those (female 1) did not show any activity towards any male. Therefore, increased parental investment by females raising chicks under uniparental care regimes results in more attractive sons.

Manipulating brood size along with sexual conflict was important because we needed to control potential parental workload. We can, however, exclude the possibility that the difference in brood size between our treatment groups were the main influence on female parental investment, because nestling begging intensity over individual feeds did not differ between nests raised under uniparental compared to biparental regimes (N.J.R., I.R.H. and G.A.P., manuscript in preparation.). Additionally, smaller broods should provide a lower overall begging stimulus to parents²⁴, yet contrary to this prediction, broods of two chicks raised by single female zebra finches received more parental investment each than did chicks in broods of four raised by pairs.

Overall, this demonstrates that sexual conflict occurs when

parents collaborate to rear young, and that parents are not entirely cooperative with one another⁵. In the conditions of our experiment, with equal potential parental workloads, full cooperation should result in equal expenditures under biparental and uniparental care. Current theory^{5,25} predicts for our conditions that sexual conflict could result in either offspring faring equally, or offspring faring worse under biparental care (Box 1). The latter can occur if the cost of parental investment is asymmetric for the sexes²⁵, or if parents respond to the parental investment of the other parent by negotiation, and show high responsiveness to any deficit by the partner⁵ (Box 1). Sexual conflict provides a clear cost to male offspring of a reduction in sexual attractiveness, and supports the view that biparental care can act to increase the number, but not the quality of offspring raised^{3,5}. This provides empirical support that post-zygotic sexual conflict is an important determinant of the strength of sexual selection², and suggests that intra-familial conflicts may be as important in determining the characteristics of life-history traits, such as egg mass and clutch size¹², as are ecological variables, such as food supply. □

Methods

Experimental protocol

Individual breeding females were put into one half of a partitioned breeding cage (120 × 45 × 40 cm), with access to an externally attached nest box. An arbitrarily selected, unrelated male was placed in the other half of the cage, but behind a partition so that neither bird could see the other. Before removal of the partition (pairing), we weighed (± 0.01 g) and measured (tarsus, wing and head–bill length ± 0.1 mm) birds. We checked nest boxes each morning, and we marked and weighed fresh eggs (± 0.01 g). At hatching, chicks were individually marked and then weighed each morning until 14 days of age. Chicks were also weighed and measured at fledging (20 days), independence (35 days), and at 50, 65 and 80 days of age. Where broods were supplemented by the addition of extra foster chicks from other nests, only original chicks (that is, full genetic siblings) were used for analyses of growth and reproduction. Unneeded offspring were fostered to other broods as necessary.

It was not always possible to have exactly four chicks in biparental-reared broods. Occasionally, broods reared one less or one more chick, but this was symmetrical with respect to group, and the overall mean was four chicks per brood. Parental investment per chick was independent of brood size, so that broods of three and five were indistinguishable from broods of four in the results. Chicks reared under uniparental and biparental regimes were of similar age (mean of brood means: uniparental, 4.86 ± 0.79 days; biparental, 4.97 ± 0.83 days) and mass (mean of brood means: uniparental, 4.02 ± 0.88 g; biparental, 3.99 ± 0.85 g) at manipulation (both $P > 0.70$; paired t -tests). Birds were freely provided with water, cuttlebone, grit, rearing food and seed daily (*ad libitum*), received a vitamin supplement and charcoal once a week, and were maintained throughout the experiment in a temperature-controlled room at 20 °C, under full spectrum, artificial light on a 16/8 h light/dark regime.

Measurement of parental investment

We established baseline feeding rates for each bird over 24 h periods for 3–4 days, by weighing seed provided and subtracting the weight of uneaten seed. Once paired, food consumption was measured daily for each pair. Parental investment was calculated, on a daily basis, as the amount of food consumed minus the parent's baseline, which was assumed to remain constant throughout the chick-rearing period. Unless otherwise stated, parental investment is expressed as the amount of parental investment per chick.

Parental provisioning effort and chick-begging behaviour were assessed using video cameras recording through a hole in the front of each nest box, which was covered when not recording. We acclimatized birds to the camera and tripod over a 24 h period before recording. However, at two nests, either one or both of the parents did not feed young during the recording session (but immediately fed the young once the camera was removed), so the sample size for analysis was 12 pairs. We measured the mass of food before and after the 3 h video recording. Chicks were weighed and individually marked before video recording, using correcting fluid, and the amount of food in each chick's crop was scored on a five-point scale (0 being empty and four being full). The total number of feeds in 3 h and the mean regurgitate duration (feed load size) for the first five feeds from each hour of video (fifteen feeds in total) were calculated for parents at all nests. Parents were scored as having fed a chick when they inserted their bill into the chick's gaping mouth and they could be seen regurgitating, with characteristic heaves of their bodies. Regurgitate duration was measured frame by frame, from the point at which parents inserted their bill until it was withdrawn from the chick's mouth.

Mate choice experiment

Twelve experienced females (all had reared at least two broods) were introduced sequentially for 10 min periods into a bi-directional mate choice cage, which had a wire mesh partition at either end, each of which was connected to a cage containing a test male. Birds could therefore see each other, but could not make physical contact. Two perches were mounted in the female compartment, which were equidistant from the cages at either end that separately housed the two males. Test males were 'dyads' of sibling males from

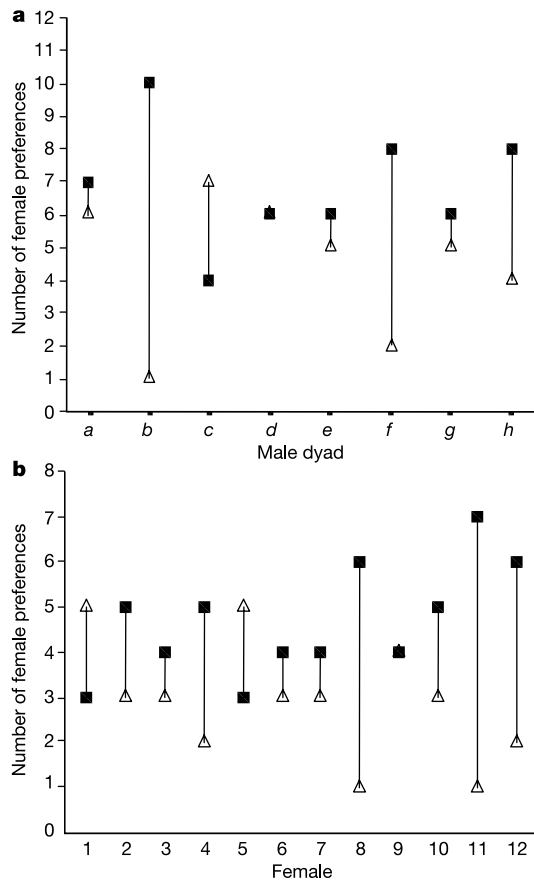


Figure 3 Mate choice experiment. **a**, Preference of females in relation to rearing regime for each male sibling-pair dyad (male dyads a–h). Females preferred males raised under the uniparental regime (Wilcoxon rank-sum test, $Z = 2.09$, $P = 0.037$). **b**, Preference of individual females in relation to rearing regime for male dyads. Again, females preferred males raised under the uniparental regime ($Z = 2.80$, $P = 0.005$). In **a** and **b**, uniparental-reared males are represented by filled squares, and biparental-reared males by triangles. Sample sizes may vary slightly between dyads owing to the loss of one female part of the way through the experiment.

uni- and biparentally reared broods. Male dyads, therefore, had the same genetic parents, and the only difference between them was that one was raised under a uniparental regime and the other a biparental regime. Males had fresh food and water available throughout the experiment. Before being introduced to the mate choice apparatus, females were housed in a cage out of sight of the test males, so were unable to assess the choice of other females or make any previous assessment of either male. The order in which females were introduced to males was varied between each sibling pair. Trials were video-recorded by N.J.R. and watched by I.R.H., who was unaware of the identity of the test males. Females were recorded as making a preference for the male on the side of the cage at which she spent most of her time during the test. The experiment was conducted under the required conditions of ultraviolet (full spectrum) lighting²⁶.

Statistical analysis

We analysed the data using SPSS for Windows 10.0 and S-Plus 2000. We tested all data for normality and homogeneity of variances before analysis. All tests are two-tailed.

Received 28 September 2001; accepted 21 January 2002.

- Clutton-Brock, T. H. *The Evolution of Parental Care* (Princeton Univ. Press, New Jersey, 1991).
- Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
- Parker, G. A. Models of parent–offspring conflict. V. Effects of the behaviour of the two parents. *Anim. Behav.* **33**, 519–533 (1985).
- Westneat, D. F. & Sargent, R. C. Sex and parenting: the effects of sexual conflict and parentage on parental strategies. *Trends Ecol. Evol.* **11**, 87–91 (1996).
- McNamara, J. M., Houston, A. I., Barta, Z. & Osorno, J.-L. Should young ever be better off with one parent than with two? *Behav. Ecol.* (in the press).
- Hurst, L. D., Atlan, A. & Bengtsson, B. O. Genetic conflicts. *Q. Rev. Biol.* **71**, 317–364 (1996).
- Godfray, H. C. J. in *Levels of Selection in Evolution* (ed. Keller, L.) 100–120 (Princeton Univ. Press, New Jersey, 1999).
- Lessells, C. M. in *Levels of Selection in Evolution* (ed. Keller, L.) 75–99 (Princeton Univ. Press, New Jersey, 1999).
- Pomiankowski, A. in *Levels of Selection in Evolution* (ed. Keller, L.) 121–152 (Princeton Univ. Press, New Jersey, 1999).
- Chapman, T. & Partridge, L. Sexual conflict as fuel for evolution. *Nature* **381**, 189–190 (1996).
- Parker, G. A. & Partridge, L. Sexual conflict and speciation. *Phil. Trans. R. Soc. Lond. B* **353**, 261–274 (1998).
- Smith, H. G. & Härdling, R. Clutch size evolution under sexual conflict enhances the stability of mating systems. *Proc. R. Soc. Lond. B* **267**, 2163–2170 (2000).
- Rice, W. R. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**, 232–234 (1996).
- Stockley, P. Sexual conflict resulting from adaptations to sperm competition. *Trends Ecol. Evol.* **12**, 154–159 (1997).
- Birkhead, T. R. & Møller, A. P. (eds) *Sperm Competition and Sexual Selection*: (Academic, New York, 1998).
- Gavrilets, S. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403**, 886–889 (2000).
- Moore, A. J., Gowaty, P. A., Wallin, W. G. & Moore, P. J. Sexual conflict and the evolution of female mate choice and male social dominance. *Proc. R. Soc. Lond. B* **268**, 517–523 (2000).
- Partridge, L. & Hurst, L. D. Sex and conflict. *Science* **281**, 2003–2008 (1998).
- Sasvári, L. Reproductive effort of widowed birds. *J. Anim. Ecol.* **55**, 553–564 (1986).
- Bart, J. & Tornes, A. Importance of monogamous male birds in determining reproductive success. *Behav. Ecol. Sociobiol.* **24**, 109–116 (1989).
- Wolf, L., Ketterson, E. D. & Nolan, V. Behavioural response of female dark-eyed juncos to experimental removal of their mates: implications for the evolution of parental care. *Anim. Behav.* **39**, 125–134 (1990).
- Whittingham, L. A., Dunn, P. O. & Robertson, R. J. Female response to reduced male parental care in birds: an experiment in tree swallows. *Ethology* **96**, 260–269 (1994).
- Markman, S., Yom-Tov, Y. & Wright, J. The effect of male removal on female parental care in the orange-tufted sunbird. *Anim. Behav.* **52**, 437–444 (1996).
- Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford Univ. Press, Oxford, 1997).
- Lessells, C. M. Parentally-biased favouritism: why should parents specialize in caring for different offspring? *Phil. Trans. R. Soc. Lond. B* (in the press).
- Bennett, A. T. D., Cuthill, I. C., Partridge, J. C. & Maier, E. J. Ultraviolet vision and mate choice in zebra finches. *Nature* **380**, 433–435 (1996).
- Smith, C. C. & Fretwell, S. D. The optimal balance between size and number of offspring. *Am. Nat.* **108**, 499–506 (1974).
- Chase, I. D. Cooperative and non-cooperative behaviour in animals. *Am. Nat.* **115**, 827–857 (1980).
- Houston, A. I. & Davies, N. B. in *Behavioural Ecology* (eds Sibley, R. M. & Smith, R. H.) 471–487 (Blackwell Scientific, Oxford, 1985).
- McNamara, J. M., Gasson, C. E. & Houston, A. I. Incorporating rules for responding into evolutionary games. *Nature* **401**, 368–371 (1999).

Acknowledgements

We thank I. P. F. Owens, L. S. Forbes, C. M. Lessells and J. M. McNamara for comments on the manuscript. The project was funded by the Natural Environment Research Council.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.R.H. (e-mail: i.hartley@lancaster.ac.uk).

Direct cortical input modulates plasticity and spiking in CA1 pyramidal neurons

Miguel Remondes & Erin M. Schuman

Caltech/Howard Hughes Medical Institute, Division of Biology, 216-76, Pasadena, California 91125, USA

The hippocampus is necessary for the acquisition and retrieval of declarative memories^{1,2}. The best-characterized sensory input to the hippocampus is the perforant path projection from layer II of entorhinal cortex (EC) to the dentate gyrus^{3,4}. Signals are then processed sequentially in the hippocampal CA fields before returning to the cortex via CA1 pyramidal neuron spikes. There is another EC input—the temporoammonic (TA) pathway—consisting of axons from layer III EC neurons that make synaptic contacts on the distal dendrites of CA1 neurons^{3,5,6}. Here we show that this pathway modulates both the plasticity and the output of the rat hippocampal formation. Bursts of TA activity can, depending on their timing, either increase or decrease the probability of Schaffer-collateral (SC)-evoked CA1 spikes. TA bursts can also significantly reduce the magnitude of synaptic potentiation at SC–CA1 synapses. The TA–CA1 synapse itself exhibits both long-term depression (LTD) and long-term potentiation (LTP). This capacity for bi-directional plasticity can, in turn, regulate the TA modulation of CA1 activity: LTP or LTD of the TA pathway either enhances or diminishes the gating of CA1 spikes and plasticity inhibition, respectively.

Using hippocampal slices optimized for stimulating both the SC and TA inputs^{7,8} (Fig. 1a, b), we examined whether TA activity can gate SC-elicited spikes in CA1 pyramidal neurons. We first began with an SC stimulus strength that consistently evoked an excitatory postsynaptic potential (EPSP) but never a spike (Fig. 1c). We found that when the SC stimulus was immediately preceded by a TA burst (10 stimuli at 100 Hz), the previously ineffective SC stimulus now evoked a spike. The spike enhancement occurred when the TA stimulus preceded the SC stimulus by 20–80 ms, suggesting temporal summation of the TA- and SC-elicited EPSPs (Fig. 1c, d). The opposite phenomenon, spike-blocking^{8,9}, can also be observed. In this case, the SC axons are stimulated at a strength that reliably elicits a CA1 spike. If a short burst is delivered to TA axons about 400 ms before the SC stimulus, SC-elicited spiking of CA1 neurons is prevented (Fig. 1e). This is due to a GABA_B (γ-aminobutyric acid)-mediated IPSP that reduces postsynaptic excitability; this IPSP can last for up to 1 s after a TA burst⁸. These data show that, depending on the relative timing of the TA input and the strength of SC stimulation, TA activity can either facilitate or block CA1 output. When a burst of TA activity precedes SC activity by ≤100 ms, SC-elicited spiking will tend to be facilitated. Conversely, TA activity that precedes SC activity by as much as 200 ms will tend to inhibit SC-elicited spiking. Because CA1 neuron activity constitutes the principal output of the hippocampal formation, these data suggest that TA activity can gate information transfer out of the hippocampus.

The capacity for TA activation to influence SC-driven spiking suggests that TA activity might also modulate plasticity at the SC–CA synapses. Indeed an earlier study indicated that stimulation of the TA pathway could reduce LTP at the SC–CA1 synapses¹⁰. We re-examined this issue in the following way: we first determined a SC theta burst stimulation (TBS) protocol that could be applied repeatedly, each time yielding roughly the same magnitude and pattern of synaptic potentiation (Fig. 2a, d). We then examined the

effects of TA activity on SC plasticity by jointly stimulating the two pathways. We used a TBS protocol to stimulate the TA path, as this most closely resembles the firing pattern of layer III entorhinal cortical neurons recorded *in vivo*^{11,12}. We found that simultaneous stimulation of the TA and SC fibres (the onset of TA TBS preceded SC TBS by 20 ms) significantly reduced the both the magnitude and the duration of SC plasticity (Fig. 2b–d). After the joint TA + SC stimulation, significantly greater plasticity was consistently elicited by a subsequent SC-alone theta burst episode (Fig. 2b–d). In addition to diminishing the magnitude of the plasticity, joint TA + SC stimulation resulted in an earlier return to baseline levels of synaptic transmission, when compared with plasticity elicited by SC stimulation alone (Fig. 2a–c). TA-induced inhibition of SC plasticity requires GABA_A-mediated transmission, as bicuculline (20 μ M) completely prevented the plasticity interference usually imposed by joint stimulation (Fig. 2d). Indeed, in the presence of bicuculline a TA-induced enhancement of plasticity was observed, indicating that, in the absence of synaptic inhibition, TA-evoked EPSPs can sum, thereby facilitating SC–CA1 plasticity. Together with a previous study¹⁰, these data demonstrate that the TA pathway can regulate the plasticity of the SC–CA1 synapse.

The gating of spikes and plasticity indicates that the TA pathway can modulate the activity of the trisynaptic circuit. However, it is also possible that the TA input provides ‘information’ as well as modulation to the hippocampus. If this is so, the capacity for plasticity of TA transmission is probably important. Although LTD of TA synaptic transmission has been observed⁷, LTP has not been

observed in hippocampal slices unless synaptic inhibition is prevented. We therefore re-examined the ability of the TA–CA1 synapses to exhibit LTP. We simultaneously recorded the SC- and TA-evoked field excitatory responses in stratum radiatum (SR) and stratum lacunosum moleculare (SLM), respectively, under conditions in which synaptic inhibition was intact. Four trains of high-frequency stimulation (HFS) to the TA axons resulted in a robust and long-lasting potentiation of synaptic transmission (Fig. 3a) (mean percentage of baseline at 60 min after tetanus, $136.1 \pm 2.9\%$; $n = 25$). The simultaneously recorded SC–CA1 response did not change upon induction of TA-LTP, providing further evidence that we obtained a complete separation of the two inputs (Fig. 3a) ($n = 25$). Consistent with the presence of NMDA (N-methyl-D-aspartate) receptors in SLM dendrites¹³ was the observation that this TA–CA1 potentiation was blocked by an NMDA-receptor antagonist, D-2-amino-5-phosphonopentanoate (AP5) (50 μ M), present during the tetanus (Fig. 3b) ($n = 7$). Thus, TA–CA1 synapses exhibit both long-lasting depression⁷ and potentiation, indicating that these synapses are capable of bi-directional synaptic modifications.

TA synaptic transmission can modulate both SC–CA1 spike activity and synaptic plasticity, and is itself subject to either upmodulation or downmodulation by LTP and LTD, respectively. Given these properties, we next examined whether the modulation of SC synaptic transmission by TA is plastic, in other words whether the TA-induced modulation of SC spikes or plasticity can be increased or decreased by LTP or LTD of the TA pathway. We addressed this in the following manner: we obtained baseline measurements of spike blocking or spike enhancement; we then induced either LTP or LTD in the TA pathway, and lastly we reassessed the efficacy of spike blocking or spike enhancement. As shown in Fig. 4, the induction of TA plasticity had a potent effect on

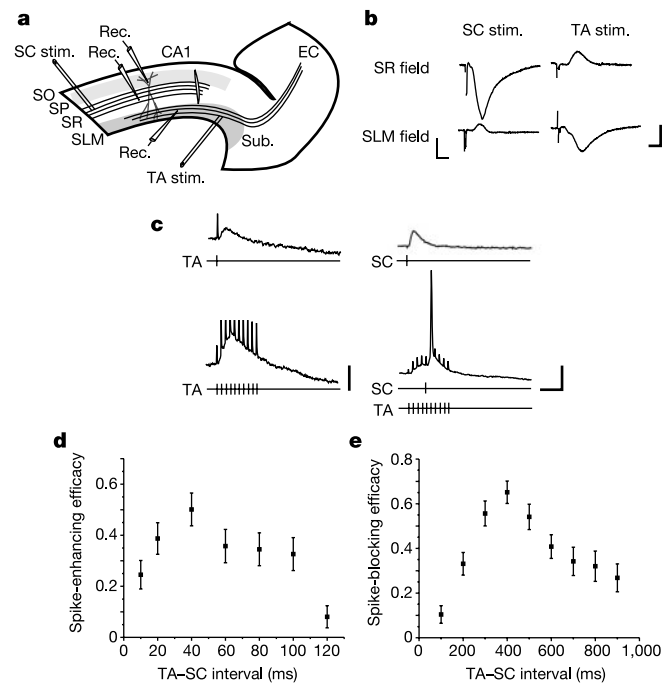


Figure 1 Enhancement and inhibition of SC-driven spikes by stimulation of TA. **a**, Recording set-up: two stimulating electrodes are used, one each in the SC and the TA. Extracellular recordings are made in SR and/or SLM. An intracellular recording from a CA1 neuron records responses elicited both in SC and in TA. **b**, Isolation of SC and TA field potentials. Stimulation of the SC results in a negative-going (sink) field potential recorded in SR and a positive-going (source) field potential in SLM. The converse is observed with TA stimulation. Scale bars, 0.5 mV (vertical), 5 ms (horizontal). **c**, Representative intracellular responses to either a single stimuli or a burst delivered to the TA or SC. When a single SC stimulus is presented alone no CA1 spike is elicited; when the SC stimulus is preceded by a TA burst that begins ~20 ms before the SC stimulus a spike is elicited. Scale bar, 5 mV (TA, vertical) or 20 mV (SC, horizontal), 50 ms. **d**, Spike enhancement as a function of the time by which the TA burst precedes the single SC stimulus ($n = 21$). **e**, Spike blocking as a function of TA–SC offset ($n = 32$).

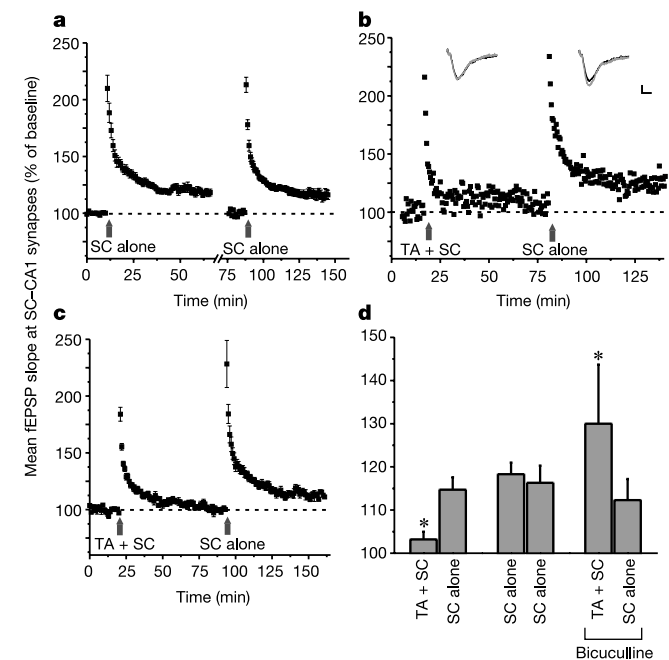


Figure 2 Interference with plasticity by stimulation of TA. **a**, Ensemble of experiments in which two consecutive epochs of TBS were delivered to the SC, each resulting in a comparable amount of potentiation (see also **d**) ($n = 12$). Dashed line, baseline synaptic transmission. **b**, When the SC TBS is coupled with TA stimulation, SC–CA1 plasticity is inhibited. Scale bars, 0.25 mV (vertical), 10 ms (horizontal). Traces are SC field before (solid) and after (dashed) stimulation. **c**, Ensemble of all plasticity interference experiments ($n = 8$). **d**, Mean SC–CA1 fEPSP potentiation, 50–60 min after stimulation. Potentiation of TA + SC was significantly lower than potentiation of SC alone. Bicuculline prevented plasticity interference and resulted in facilitation ($n = 6$).

the CA1 spike gating. To avoid ceiling or floor effects, the initial magnitude of either spike enhancement or blocking determined whether TA-LTP or TA-LTD would be subsequently induced. In experiments in which little spike enhancement was initially observed, the induction of TA-LTP greatly facilitated spike enhancement (Fig. 4a, b). Conversely, when strong spike enhancement was initially observed, its magnitude was markedly reduced by the induction of LTD at TA–CA1 synapses (Fig. 4c, d). The spike blocking phenomenon was also potently modulated by TA plasticity. In experiments in which low levels of spike blocking were initially observed, TA-LTP facilitated subsequent spike-blocking attempts (Fig. 4e, f). Conversely, after TA-LTD the spike-blocking efficacy was reduced (Fig. 4g, h). Thus, the effect of TA plasticity was either to increase or to decrease the gain of spike gating: LTP increased the magnitude of both spike enhancement and blocking, whereas LTD decreased both phenomena.

We next determined whether plasticity in the TA pathway also changes its ability to modulate SC–CA plasticity. We assessed the potency of plasticity interference before or after establishing TA plasticity. As before, the initial level of plasticity interference dictated whether TA-LTP or LTD would be induced: high initial levels of plasticity interference were usually followed by LTD induction, whereas low levels were usually followed by LTP induction. The induction of LTP at TA synapses consistently increased the magnitude of plasticity interference (Fig. 5a, b); joint TBS of the TA and SC pathway resulted in smaller amounts of potentiation at the SC–CA1 synapses after TA-LTP (see Supplementary Information for simultaneously recorded TA responses). Thus, LTP induction at the TA synapses increased its ability to interfere with SC–CA1 plasticity. The opposite was true of LTD induction: interference with plasticity was reduced. The magnitude of SC–CA1 potentiation observed after TA-LTD was greater, indicating that the modulatory capabilities of the TA pathway were reduced (Fig. 5c, d). A change in the degree of plasticity interference was observed only when TA plasticity was induced; in separate experiments we examined the stability of plasticity interference by repeatedly testing the magnitude of potentiation observed after joint TA + SC stimulation. When plasticity interference was initially low it stayed low in the absence of plasticity (Fig. 5b); conversely, when plasticity interference was initially high it stayed high in the absence of TA plasticity (Fig. 5d). Thus, like the spike gating imposed by TA activity, the plasticity interference is also subject to significant modulation by plasticity of the TA–CA1 synapses themselves.

Thus, our data show that the TA–CA1 synapses can be either potentiated or depressed; this modification of the TA synapses by plasticity in turn modulates this pathway's ability to gate the output

or plasticity of SC–CA1 synapses. Modulation of excitatory synaptic connections on inhibitory interneurons whose processes extend into SLM also influences the strength of TA–CA1 activity^{8,14}. The results indicate that the net TA modulation of CA3–CA1 synaptic transmission will be influenced dynamically by both the strength of the TA–CA1 synaptic connections as well as the relative timing of the TA and trisynaptic inputs to CA1. Detailed knowledge of the temporal firing properties of the different entorhinal cortical layers is needed for a full understanding of the interplay between these two inputs to CA1 pyramidal neurons.

These results indicate an important modulatory role for the TA pathway but also raise the possibility of a more fundamental role of TA–CA1 synaptic transmission in hippocampal information processing. The hippocampal formation is firmly established as being necessary for the formation and retention of declarative memories. In characterizing the synaptic contributions that underlie hippocampus-dependent memories, most studies have focused on entorhinal input to the dentate gyrus that is then serially processed

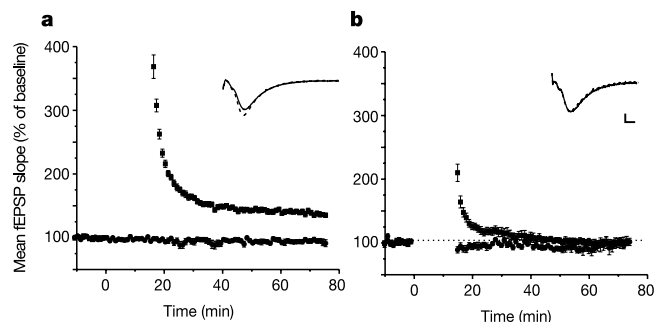


Figure 3 TA–CA1 synapses exhibit NMDA-receptor-dependent LTP. **a**, Ensemble of all TA LTP experiments. HFS resulted in significant potentiation of TA–CA1 synaptic transmission (squares) without affecting the SC–CA1 synaptic responses (circles) recorded simultaneously. **b**, Ensemble of all TA-LTP experiments conducted in the presence of an NMDA-receptor antagonist, AP5. Scale bars in insets, 0.1 mV (vertical), 10 ms (horizontal).

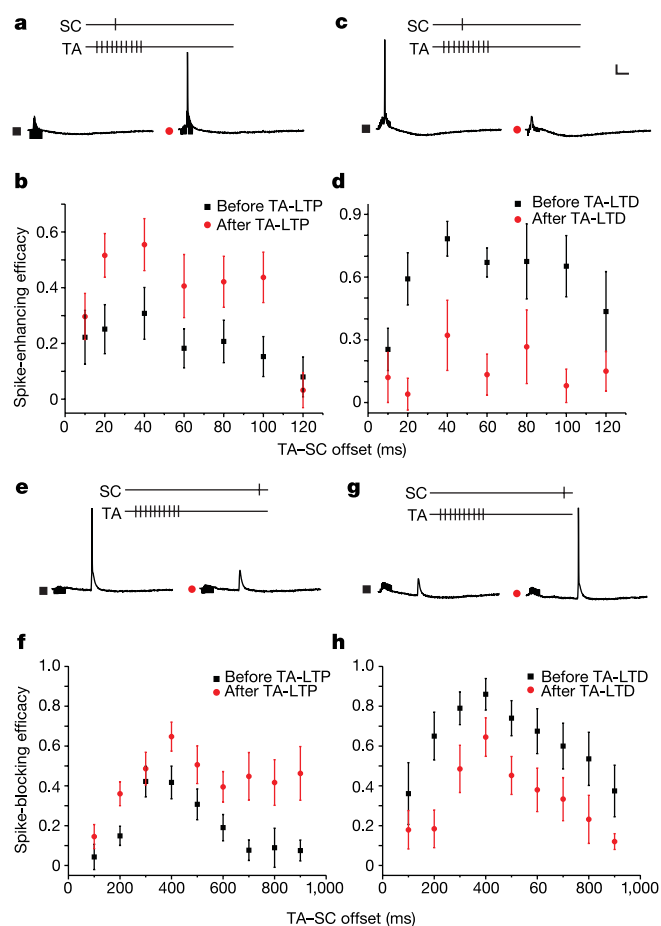


Figure 4 Modulation of spike blocking and spike enhancement by TA-LTP and LTD. **a**, Spike-enhancement protocol and CA1 neuron intracellular record before (black square) and after (red circle) TA-LTP. **b**, Summary of spike enhancement (TA–SC offset is the time by which the TA burst precedes the SC stimulus) before and after TA-LTP ($n = 15$). **c**, Spike enhancement protocol and CA1 neuron intracellular record before (black square) and after (red circle) TA-LTD. **d**, Analysis of all experiments showing the amount of spike enhancement achieved before and after TA-LTD ($n = 6$). **e**, Spike-blocking protocol and CA1 neuron record before (black square) and after (red circle) TA-LTP. **f**, Summary data showing spike blocking before and after TA-LTP ($n = 22$). **g**, Spike-blocking protocol and CA1 neuron record before (black square) and after (red circle) TA-LTD. **h**, Summary of spike blocking before and after TA-LTD ($n = 10$). Scale bars for **a**, **c**, **e**, **g**: 10 mV (vertical), 100 ms (horizontal).

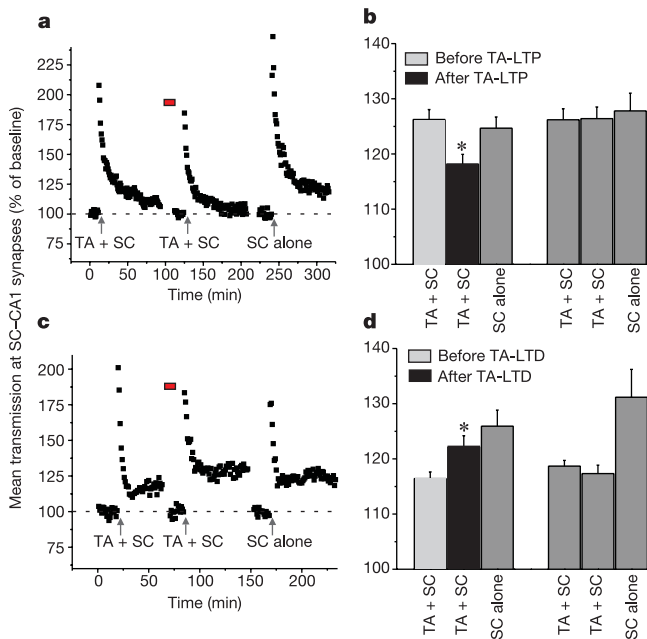


Figure 5 Modulation of interference with plasticity by TA-LTP and LTD. **a**, Representative experiment monitoring SC-CA1 transmission in which LTP in the TA pathway (at the time indicated by the red-filled oblong) increased the plasticity interference. **b**, Summary of the potentiation for each TA + SC stimulation epoch (first two bars) before and after LTP and the enhancement observed after SC stimulation alone. TA-LTP significantly increased plasticity interference ($n = 24$). In separate control experiments (right bars), no change in plasticity interference occurred when the TA + SC stimulation was applied again ($n = 5$). **c**, Representative experiment in which TA-LTD (at the time indicated by the red-filled oblong) decreased plasticity interference. **d**, Summary of the potentiation for each TA + SC stimulation epoch before and after LTD and the enhancement observed after SC stimulation alone. TA-LTD significantly decreased plasticity interference ($n = 11$). In separate control experiments no change in plasticity interference occurred when TA + SC stimulation was applied again ($n = 5$).

by synapses in areas CA3 and CA1 (the trisynaptic circuit). However, some studies suggest that the trisynaptic circuit is not necessary for some kinds of spatial memory^{15,16}. A distinguishing feature of pyramidal neurons of area CA3 or CA1 is their firing selectivity for particular locations in the animal's environment, the so-called 'place fields'¹⁷. These place fields are thought to contribute to hippocampus-dependent spatial behaviours. Some studies have shown that the trisynaptic circuit is not required for the establishment of place-selective cells^{18,19}. In these studies, the ablation of dentate granule cells¹⁸ or CA3 pyramidal neurons¹⁹ does not abolish CA1 place fields. Our plasticity results suggest that, in addition to modulating trisynaptic hippocampal function, the TA pathway could provide the information required for important aspects of memory and place-cell function. □

Methods

Slice preparation and electrophysiology

Slices were prepared and microdissected to isolate the TA pathway, as described previously^{7,8}. In brief, after decapitation and removal of the brain, the dorsal surface of the posterior half of each hemisphere was glued to the stage of a cooled oscillating tissue slicer and covered with chilled artificial cerebrospinal fluid (ACSF) containing (in mM) 119 NaCl, 2.5 KCl, 1.3 MgSO₄, 2.4 CaCl₂, 1.0 NaH₂PO₄, 26.3 NaHCO₃, 11.0 glucose. Slices (500 μ m) were cut, with the optimal slices (as assessed visually, by ease of identification of distinct layers, as well as electrophysiologically, by the presence of robust field potentials) generally found 4–4.5 mm below the ventral surface. Bipolar tungsten electrodes, either concentric or paired needles, were used for stimulation. One electrode was placed in SR to stimulate the SCs; the other in SLM to stimulate the TA pathway. Unless otherwise noted, all experiments were conducted with synaptic inhibition intact. Field recordings were made with low-resistance micropipettes filled with 3 M NaCl. The SC response was recorded in SR and the TA response in SLM. To clearly isolate the TA response it was necessary to dissect the slice further^{7,8}. The entire dentate gyrus was removed to eliminate

the possibility of activation of the trisynaptic pathway and to prevent the contamination of a field response recorded in SLM by the much larger field elicited in the dentate gyrus by concurrent activation of the perforant path. In most experiments CA3 was also removed to eliminate the possibility of disinaptic activation via the perforant path projection to CA3. In addition, a cut was made through SR in distal CA1 (near the subiculum) perpendicular to the cell body layer to prevent the antidromic activation of SC by the stimulation electrode in SLM. SCs do not enter SLM to any appreciable extent⁴, so this cut cleanly isolates TA axons. Separation of the two pathways was confirmed by the observation of a positive-going field potential in the other layer¹⁵. Intracellular recording was conducted as described previously^{7,8}. The following stimulation protocols were used: HFS, 100 Hz for 1 s, repeated 4 times at 5-min intervals; TBS, 8 bursts of either 5 (SC) or 10 (TA) pulses at 100 Hz, 200 ms between bursts; low-frequency stimulation (LFS), stimulation at 1 Hz for 10 min. All stimulus pulses were of the same length and amplitude as test pulses. Test pulses were applied once every 30 s to each pathway. The initial slope of the field EPSP was measured. Drugs were applied by dilution of concentrated stock solutions into the perfusion medium. All chemicals were obtained from Sigma (St Louis, Missouri).

Data analysis

Data were collected directly by an IBM-compatible computer using in-house software. All numerical values are listed as means $\pm$ s.e.m. unless otherwise stated. Depression and potentiation were measured by taking an average of the initial slopes of the field EPSPs (fEPSPs) over a 10-min period before and after the end of LFS, HFS or TBS. Spike-enhancing efficacy was calculated by measuring the spike probability (p) as follows: $[p(\text{TA} + \text{SC}) - p(\text{SC alone})]/[1 + p(\text{SC alone})]$.

Spike-blocking efficacy was calculated by measuring the spike probability (p) as follows: $[p(\text{SC alone}) - p(\text{TA} + \text{SC})]/p(\text{SC alone})$. Student's paired t -test was used to determine statistical significance for within-group comparisons; the unpaired t -test was used between groups. Data reported as significant (indicated by an asterisk in figures) achieved at least the $P \leq 0.01$ level. Points in figures represent means $\pm$ s.e.m. across all experiments. Representative traces, shown in insets, are averages of five consecutive sweeps from a representative experiment, taken 5 min before LFS, HFS or TBS, and 60 min after the end of LFS, HFS or TBS.

Received 26 December 2001; accepted 28 January 2002.

- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nature Rev. Neurosci.* **1**, 41–50 (2001).
- Witter, M. P., Groenewegen, H. J., Silva, F. H. L. D. & Lohman, A. H. Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Prog. Neurobiol.* **33**, 161–253 (1989).
- Amaral, D. G. & Witter, M. P. The three-dimensional organisation of the hippocampus: a review of the anatomical data. *Neuroscience* **31**, 571–591 (1989).
- Cajal, S. R. *The Structure of Ammon's Horn* (ed. Thomas, C. C.) (Thomas, Springfield, Illinois, 1968).
- Steward, O. & Scoville, S. A. Cells of origin of entorhinal cortical afferents to the hippocampus and fascia dentate of the rat. *J. Comp. Neurol.* **169**, 347–370 (1976).
- Dvorak-Carbone, H. & Schuman, E. M. Long-term depression of temporoammonic-CA1 hippocampal synaptic transmission. *J. Neurophysiol.* **81**, 1036–1044 (1999).
- Dvorak-Carbone, H. & Schuman, E. M. Patterned activity in stratum lacunosum moleculare inhibits CA1 pyramidal neuron firing. *J. Neurophysiol.* **82**, 3213–3222 (1999).
- Empson, R. M. & Heineman, U. The perforant path projection to hippocampal area CA1 in the rat hippocampal-entorhinal cortex combined slice. *J. Physiol. (Lond.)* **484**, 707–720 (1995).
- Levy, W. B., Desmond, N. L. & Zhang, D. X. Perforant path activation modulates the induction of long-term potentiation of the Schaffer collateral-hippocampal CA1 response: theoretical and experimental analyses. *Learn. Mem.* **4**, 510–518 (1998).
- Chrobak, J. J. & Buzsaki, G. Gamma oscillations in the entorhinal cortex of the freely behaving rat. *J. Neurosci.* **18**, 388–398 (1998).
- Chrobak, J. J., Lorincz, A. & Buzsaki, G. Physiological patterns in the hippocampo-entorhinal cortex system. *Hippocampus* **10**, 457–465 (2000).
- Colbert, C. M. & Levy, W. B. Electrophysiological and pharmacological characterization of perforant path synapses in CA1: mediation by glutamate receptors. *J. Neurophysiol.* **68**, 1–8 (1992).
- Maccaferri, G. & McBain, C. J. Passive propagation of LTD to stratum oriens-alveus inhibitory neurons modulates the temporoammonic input to the hippocampal CA1 region. *Neuron* **15**, 137–145 (1995).
- Jarrard, L. E., Okaichi, H., Steward, O. & Goldschmidt, R. B. On the role of hippocampal connections in the performance of place and cue tasks: comparisons with damage to hippocampus. *Behav. Neurosci.* **98**, 946–954 (1984).
- Jarrard, L. E. What does the hippocampus really do? *Behav. Brain Res.* **71**, 1–10 (1995).
- Muller, R. A quarter of a century of place cells. *Neuron* **17**, 813–822 (1996).
- McNaughton, B. L., Barnes, C. A., Meltzer, J. & Sutherland, R. J. Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Exp. Brain Res.* **76**, 485–496 (1989).
- Brun, V. H., Otnass, M. K., Witter, M. P., Moser, M. B. & Moser, E. I. Place representation in hippocampal area CA1 in the absence of input from area CA3. *Soc. Neurosci. Abstr.* **31** (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank G. Laurent for critical reading of the manuscript. This work was supported by Fundacao para a Ciencia e Tecnologia (FCT)—Portugal and the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.M.S. (e-mail: schumane@its.caltech.edu).

Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation

Eliana Drenkard & Frederick M. Ausubel

Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Colonization of the lungs of cystic fibrosis (CF) patients by the opportunistic bacterial pathogen *Pseudomonas aeruginosa* is the principal cause of mortality in CF populations^{1,2}. *Pseudomonas aeruginosa* infections generally persist despite the use of long-term antibiotic therapy^{1,3}. This has been explained by postulating that *P. aeruginosa* forms an antibiotic-resistant biofilm^{4,5} consisting of bacterial communities embedded in an exopolysaccharide matrix. Alternatively, it has been proposed that resistant *P. aeruginosa* variants may be selected in the CF respiratory tract by antimicrobial therapy itself^{1,6}. Here we report that both explanations are correct, and are interrelated. We found that antibiotic-resistant phenotypic variants of *P. aeruginosa* with enhanced ability to form biofilms arise at high frequency both *in vitro* and in the lungs of CF patients. We also identified a regulatory protein (PvrR) that controls the conversion between antibiotic-resistant and antibiotic-susceptible forms. Compounds that affect PvrR function could have an important role in the treatment of CF infections.

Antibiotic-resistant colonies of *P. aeruginosa* clinical isolate PA14 (ref. 7) arose at a frequency of 10^{-6} to 10^{-7} when cultures were plated on Luria-Bertani (LB) agar containing kanamycin ($200 \mu\text{g ml}^{-1}$). Resistant colonies were smaller than wild-type even on antibiotic-free media, and exhibited colony morphologies similar to the ones described for CF variant isolates^{6,8}. One class of the resistant variants (approximately 30%) exhibited a rough colony phenotype compared with the wild type, and was called RSCV (rough small-colony variant). When RSCVs were grown on antibiotic-free LB agar, wild-type revertants, characterized by a large colony size, smooth appearance, and wild-type levels of susceptibility to kanamycin, arose on the edges of the variant colonies after five days incubation at room temperature (Fig. 1a), suggesting that the phenotypic changes observed in the resistant variants were transient. In addition to being resistant to kanamycin, (40 times the wild-type susceptibility level), individual RSCV colonies were also resistant to amikacin ($30 \mu\text{g ml}^{-1}$), carbenicillin ($300 \mu\text{g ml}^{-1}$), gentamicin ($30 \mu\text{g ml}^{-1}$), tobramycin ($10 \mu\text{g ml}^{-1}$) and tetracycline ($150 \mu\text{g ml}^{-1}$). Consistent with this latter result, resistant variants were also obtained at frequencies of about 10^{-7} by plating cultures of PA14 on media containing similar concentrations of the antibiotics mentioned above.

Although RSCV colonies were smaller than wild type, their small colony size was not a consequence of slow growth, as the generation time of RSCV in liquid medium was not significantly different from that of the wild type, even in LB agar containing $200 \mu\text{g ml}^{-1}$ kanamycin. Unlike the wild type, RSCV formed visible aggregates when liquid cultures were left without shaking at room temperature

(Fig. 1b). Moreover, RSCVs exhibited increased attachment to glass (not shown) and polyvinylchloride plastic (PVC) (Fig. 1c). Reverted RSCV showed wild-type levels of both agglutination and attachment to glass and PVC plastic (data not shown).

Because the ability of bacteria to attach to each other and to surfaces depends in part on the interaction of hydrophobic domains⁹, we determined the surface hydrophobicity of RSCV relative to the wild type. RSCV clones agglutinated at a lower salt concentration (0.125 M) than wild-type PA14 (0.5 M), indicating that RSCVs had a higher degree of surface hydrophobicity. Similarly, 200 mM tetramethyl urea (TMU), a hydrophobic bond-breaking agent, reduced the attachment of RSCV cells to PVC plates to wild-type levels (data not shown). Several RSCV clones were tested in the experiments described above, and all exhibited similar phenotypes. A single RSCV clone, referred to as RSCV for simplicity, was therefore chosen for further analysis.

Recently, it has been shown that *P. aeruginosa* in the sputum of CF patients exists primarily as a biofilm⁴. To determine whether the antibiotic-resistant phenotype of RSCV is associated with altered biofilm formation, biofilms of PA14 or RSCV expressing green fluorescent protein (GFP) were cultivated in flow chambers under continuous culture conditions. Analysis of biofilm structures using confocal scanning laser microscopy (CSLM) showed that RSCV formed biofilm faster (RSCV microcolonies appeared 4–5 h earlier

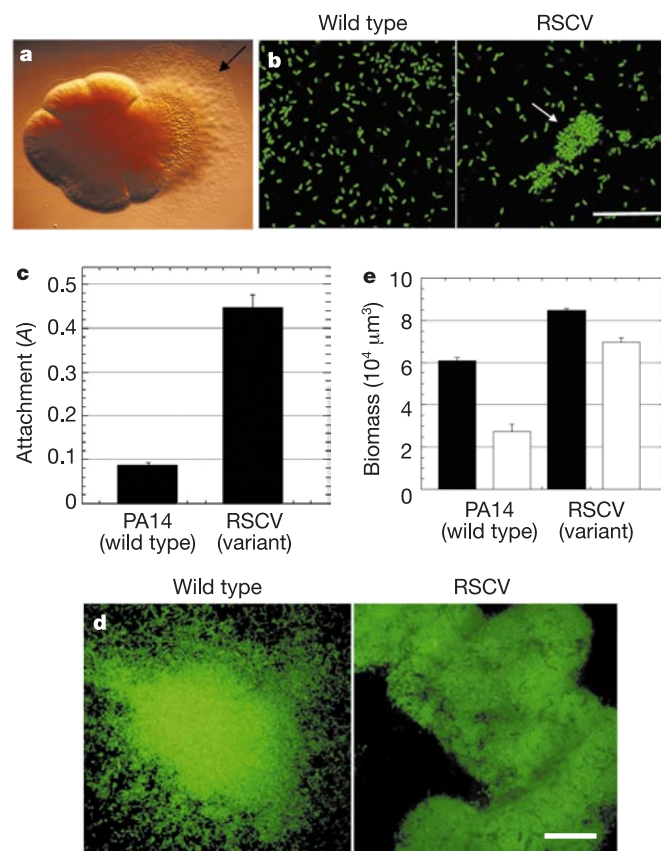


Figure 1 Characterization of *P. aeruginosa* antibiotic-resistant RSCV. **a**, Reversion of RSCV was observed on the edges of colonies (arrow) on antibiotic-free media, $30 \times$. **b**, CSLM analysis of bacterial aggregates (arrows) expressing GFP after growth in liquid broth. Scale bar, $25 \mu\text{m}$. **c**, Attachment of wild-type PA14 and RSCV to PVC (see Methods). **d**, Attachment of wild-type PA14 and RSCV to PVC (see Methods). **e**, Attachment of wild-type PA14 and RSCV to PVC (see Methods). **d**, CSLM analysis of biofilm formed by wild-type PA14 and RSCV expressing GFP. Scale bar, $50 \mu\text{m}$. **e**, Resistance of biofilms to tobramycin was determined by measuring viable biomass on 45-h-old established biofilms before (filled bars) and after (open bars) 36-h tobramycin ($200 \mu\text{g ml}^{-1}$) treatment. In the absence of antibiotic treatment, both PA14 and RSCV biofilms grew at constant rates (not shown).

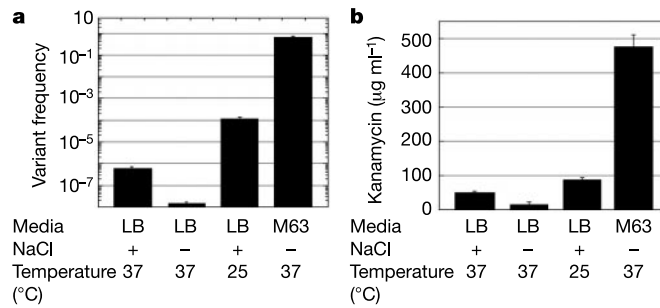


Figure 2 Appearance of phenotypic variants resistant to kanamycin depends on environmental factors. **a**, Effect of different environmental stimuli on the rate of appearance of antibiotic-resistant variants on different media containing 200 μg ml⁻¹

kanamycin (see Methods). **b**, Minimal inhibitory concentrations of kanamycin for strain PA14 using different conditions (MICs were determined as described previously²⁵).

than wild type) and had more biomass than wild type (Fig. 1d). Moreover, wild type and RSCV displayed different patterns of biofilm development. After 21 h, PA14 wild-type formed flat circular microcolonies, whereas RSCV formed irregularly shaped three-dimensional structures that did not show the typical *P. aeruginosa* biofilm morphology (Fig. 1d). Additionally, measurements of viable biomass of GFP-tagged PA14 and RSCV cells using CSLM analysis¹⁰ showed that biofilms formed by RSCV were more resistant to a continuous flow of tobramycin (200 μg ml⁻¹) than wild-type PA14 biofilms (Fig. 1e), paralleling the resistance observed on plates.

Phenotypic (phase) variation is a common phenomenon in Gram-negative bacteria that often involves environmentally regulated changes in surface components leading to alterations in observable phenotypes¹¹. We therefore examined the effect that different environmental stimuli had on the appearance of kanamycin-resistant phenotypic variants. As shown in Fig. 2a, we observed a 40-fold increase in the frequency of appearance of resistant variants (not just RSCV) obtained on LB media containing NaCl (85 mM) compared with the same medium without NaCl. Moreover, the frequency of variants increased 200-fold when plates were incubated at 25 °C compared with 37 °C (Fig. 2a). Finally, a marked 10⁶-fold increase was obtained on minimal M63 salts compared with LB medium (Fig. 2a). Notably, there was a correlation between the frequency of appearance of kanamycin-resistant variants on plates and minimal inhibitory concentrations (MICs) of kanamycin in liquid culture (Fig. 2b). For example, the high frequency of resistant variants obtained on M63 plates correlated with the relatively high concentration of kanamycin (475 μg ml⁻¹) required to inhibit the growth of PA14 in M63 liquid medium (Fig. 2a, b). These data suggest that the components involved in the formation of antibiotic-resistant variants are regulated by environmental signals. Moreover, the data indicate that the portion of the population that becomes resistant to antibiotics through phenotypic variation is dependent on environmental conditions.

To investigate whether antibiotic treatment in *P. aeruginosa* CF infections selects for resistant variants, we looked for the presence of small-colony variants in CF sputum samples. The analysis of five CF

sputum samples obtained from the Clinical Microbiology Laboratory at Massachusetts General Hospital revealed that two of the samples (5 and 38) contained 100% small-colony variants (Table 1) that reverted to wild type on prolonged incubation on antibiotic-free medium. Importantly, both samples 5 and 38 corresponded to patients that were undergoing antibiotic treatment at the time the samples were obtained (intravenous amikacin/ceftazidime for two days, and oral levofloxacin/inhaled tobramycin for six weeks, respectively; Table 1). Moreover, there was 29% enrichment in small-colony variants in samples taken on two consecutive days from the patient that was undergoing intravenous antibiotic treatment. As shown in Table 1, 30–100% of the small-colony variants present in samples 5 and 38 were resistant to four different antibiotics (amikacin, gentamicin, tetracycline and tobramycin) at concentrations equal to or higher than the minimal bactericidal concentration (MBC) of their respective reverted colonies. Although the three other CF sputum samples (41, 42 and 43) appeared to contain either a small proportion or no detectable small-colony variants when plated on antibiotic-free media, they did contain a considerable number (0.5–15%) of antibiotic-resistant variants (Table 1). This discrepancy was due to the fact that it took the small-colony variants 36–40 h to form visible colonies, at which time the fast-growing wild-type bacteria present in the sputum samples had overgrown the antibiotic-free plates.

Previous reports have shown that phenotypic switching in *Pseudomonas tolaasii* is modulated by an environmentally responsive regulatory factor¹². To identify putative *P. aeruginosa* factor(s) involved in the switching between antibiotic resistant and susceptible forms, we transferred a cosmid library of PA14 chromosomal DNA⁷ en masse to RSCV, and screened 2,500 transconjugants for colonies displaying wild-type colony size and morphology on antibiotic-free LB medium. We identified a cosmid clone (pED20) that was capable of greatly enhancing the rate of RSCV reversion. Whereas RSCV colonies were normally very stable in liquid culture (that is, no revertants were observed after 12 h incubation of an overnight culture plated on LB agar), 100% of the pED20-carrying RSCV cells formed wild-type colonies on LB plates after overnight incubation. A subclone of pED20, pED202, which contained a 3.5-

Table 1 Occurrence of phenotypic variation in *P. aeruginosa* CF isolates

	Sputum samples				
	Sample 5	Sample 38	Sample 41	Sample 42	Sample 43
Antibiotic treatment of CF patients	Amikacin (IV), ceftazidime (IV)	Tobramycin (I), levofloxacin (O)	None	None	None
Small-colony variants in sample (%)	100	100	<0.11	0	<0.12
Variants resistant to amikacin (%)	100	100	15	5	0.2
Variants resistant to gentamicin (%)	100	100	10	6.6	0.5
Variants resistant to tetracycline (%)	30	32	0	0	ND
Variants resistant to tobramycin (%)	50	100	0.10	0	0.5

The modes of antibiotic administration in CF patients were: intravenous (IV), inhalation (I), or oral (O). ND, not determined.

kilobase (kb) fragment, restored the attachment (Fig. 3a), colony morphology and auto-agglutination phenotypes (data not shown) of RSCV cells to wild-type.

The pED202 clone contains a single open reading frame (ORF) that shows sequence similarities to response regulator elements of two-component regulatory systems, including 30% identity and 45% similarity to the *Vibrio cholerae* response regulator VieA. The ORF on pED202 (designated *pvrR* for phenotype variant regulator) also shows 29% identity and 45% similarity to a probable two-component response regulator identified in *P. aeruginosa* strain PAO1 (ref. 13) (ORF PA3947). A homology search against domain sequences (ProDom; <http://prodes.toulouse.inra.fr/prodom/doc/prodom.html>) identified four regions with high-scoring segment pairs in PvrR (Fig. 3b). All four domains are also present in VieA and the PAO1 putative regulator (Fig. 3b). Moreover, these four domains exhibit high levels of amino-acid sequence similarity (30–60%; Fig. 3b). Sequence analysis of the regions located upstream and downstream of *pvrR* revealed the presence of two additional ORFs (designated ORF1 and ORF3; Fig. 3c) with sequence homology to two-component regulatory elements. The protein coded for by ORF1 has homology to probable sensor/response regulator hybrids from *P. aeruginosa* (33% identity and 46% similarity to ORF PA2824), and the protein coded for by ORF3 shows 46% identity and 63% similarity to the GacS sensor kinase from *P. fluorescens*. To determine whether *pvrR* or a highly similar *pvrR* homologue is present in other *P. aeruginosa* strains, we performed polymerase chain reaction (PCR) analysis of 14 *P. aeruginosa* strains using PvrR-specific primers. We subsequently confirmed the specificity of the PCR products obtained by Southern blotting and hybridization with a *pvrR*-specific probe. Seven out of seven CF isolates, two out of three clinical isolates and three out of four standard *P. aeruginosa* laboratory strains contained the *pvrR* gene fragment or a highly similar fragment (data not shown).

Consistent with the putative role of PvrR in the regulation of phenotypic switching, overexpression of PvrR from the pED202 clone in wild-type PA14 resulted in a sixfold reduction in the frequency of resistant variants obtained after plating overnight

cultures on kanamycin plates ($200 \mu\text{g ml}^{-1}$) compared with wild-type (Fig. 3d). Interestingly, the PvrR-overexpressing strain also caused a 2.5-fold reduction in attachment to PVC plastic with respect to the strain carrying the vector alone (Fig. 3e).

As PvrR is involved in the regulation of the phenotypic switch, we postulated that mutation of *pvrR* would alter the proportion of resistant variants present in the PA14 population. Indeed, a 914-bp in-frame deletion of *pvrR* ($\Delta pvrR$) in PA14 exhibited increased frequency of appearance of resistant variants on kanamycin plates with respect to the wild type (Fig. 3f), confirming the involvement of *pvrR* in the regulation of phenotypic switching. However, because 100% of the variants expressing *pvrR* reverted to the wild-type phenotype, we propose that PvrR is implicated primarily in inducing reversion from variant to wild-type phenotypes. In addition, as PvrR overexpression reverts RSCV with 100% efficiency, it is unlikely that RSCV formation is mechanistically related to the high frequency hypermutability observed in *P. aeruginosa* CF isolates¹⁴. Finally, our results suggest that PvrR may be acting upstream of the phenotypic switch, as inactivation of *pvrR* by mutation does not result in conversion to the variant type.

The data presented in here indicate that *P. aeruginosa* is capable of undergoing transient phenotypic changes that allow the bacteria to increase their antibiotic resistance both *in vitro* and *in vivo*. Analogous to phase variation phenomena, antibiotic-resistant RSCV cells representing a relatively small fraction of the bacterial population ensure its survival in the presence of antibiotics. It remains to be determined whether the underlying mechanism of RSCV formation involves the types of DNA rearrangements that characterize other phase variable systems^{11,15}.

Because some of the phenotypes found in PA14 RSCV have been associated with the emergence of antibiotic resistance in bacterial biofilms^{16,17}, we speculate that resistant phenotypic variants present in *P. aeruginosa* biofilms are responsible for the increased resistance to antimicrobial agents observed in CF infections by *P. aeruginosa*. Moreover, we propose that variant phenotypes selected inside mature biofilms by antibiotic treatment and other conditions present in the lung of CF patients or in the biofilm itself (such as

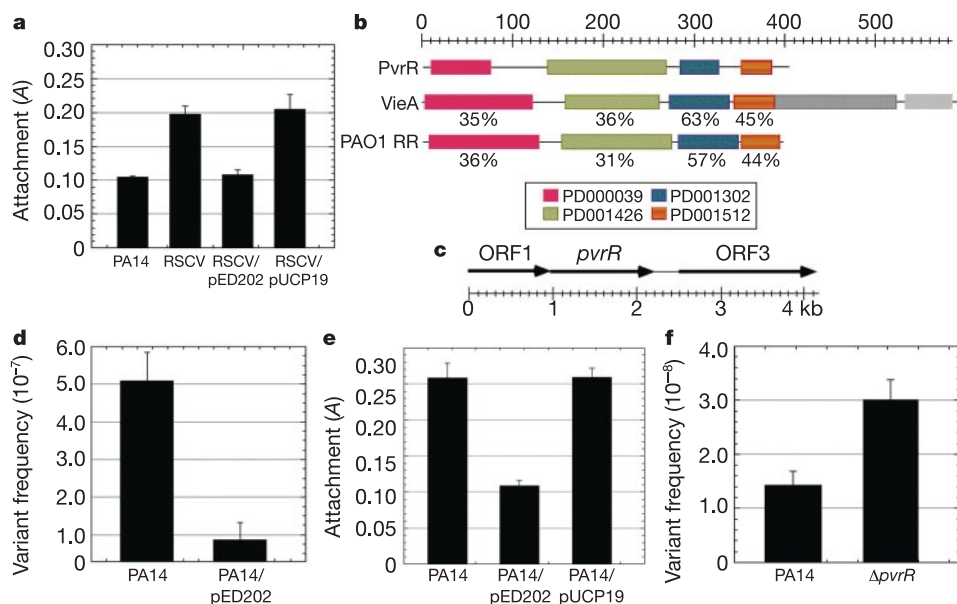


Figure 3 Identification and characterization of the *P. aeruginosa* PvrR response regulator. **a**, Attachment of strains to PVC plastic (see Methods). **b**, Alignment of PvrR-predicted amino-acid sequence with *V. cholerae* VieA and the *P. aeruginosa* PAO1 putative response regulator PA3947 (PAO1 RR). Numbers above scale indicate number of amino acids. The lower panel contains domain family numbers according to ProDom

nomenclature. **c**, The *pvrR* gene is flanked by ORF1 and ORF3 with the same transcriptional orientation. **d**, Effect of PvrR overexpression on the frequency of variant formation. **e**, Effect of PvrR overexpression on attachment to PVC. **f**, Deletion of *pvrR* affects the frequency of variants.

nutrient limitation) constitute the so-called resistant biofilm phenotype. Identification of a two-component response regulator, PvrR, that modulates the phenotypic switch from antibiotic-resistant to antibiotic-susceptible forms and also regulates biofilm formation, reinforces the possibility that these mechanisms are interrelated. Importantly, our data suggest that activation of regulatory elements like PvrR could prevent biofilm formation and render bacteria more susceptible to antimicrobial agents. Because the appearance of phenotypic variants in response to antibiotic treatment has been reported in both Gram-negative and Gram-positive bacteria¹⁸, resistance mechanisms similar to the one found in this study may be common among other bacterial pathogens. □

Methods

Bacterial strains, media and culture conditions

Clinical and CF *P. aeruginosa* isolates were provided by G. Pier or obtained from the Massachusetts General Hospital Clinical Microbiology Laboratory. M63 medium¹⁹ was supplemented with 0.3% glucose, 1 mM MgSO₄ and 0.5% casamino acids (CAA). Modified FAB medium (0.1 mM CaCl₂, 0.01 mM Fe-EDTA, 0.15 mM (NH₄)₂SO₄, 0.33 mM Na₂HPO₄, 0.2 mM KH₂PO₄ and 1 mM MgCl₂) contained 0.5% CAA and 10 mM sodium citrate. Antibiotic concentrations (per ml) were: 100 µg ampicillin and 30 µg kanamycin for *Escherichia coli*; 300 µg carbenicillin for *P. aeruginosa*. We purchased antibiotics from Sigma.

Attachment and cell surface hydrophobicity assays

Attachment assays²⁰ were performed in 96-well PVC plates using M63 supplemented with CAA (0.5%), MgSO₄ (1 mM) and glucose (0.3%) or LB media. We determined hydrophobic surface properties using the salt aggregation test (SAT)²¹.

Continuous-culture biofilm

Bacteria were cultivated in M63 in flow chambers with channel dimensions of 12 × 52 × 2 mm (provided by M. Franklin) for biofilm characterization, or in FAB medium in flow chambers with channel dimensions of 1 × 40 × 4 mm (Stovall) for measurement of biofilm resistance. The flow system was prepared and assembled as described²². Flow cells in both cases were inoculated with 100-fold dilutions of overnight cultures of PA14 or RSCV expressing GFP on plasmid pSMC21 (derivative of pSMC2 (ref. 23) provided by G. O'Toole). After inoculation, the medium flow was stopped for 1 h. Medium flow was resumed at a rate of 0.21 ml min⁻¹ using a peristaltic pump (IsmaTec), and the flow-cell system was incubated at 37 °C. We used a Leica TCS SP system (Leica Lasertechnik; GmH) for CSLM. Image analysis of antibiotic-treated biofilms was done in structures contained within serial section stacks of images delimited by freehand drawing. Pixel intensities unique to GFP-labelled bacteria and surrounding biofilm were established by the threshold limit technique. The volume (in µm³) of individual biofilm structures was determined from serial sections using ImageSpace software (Molecular Dynamics).

Frequency of appearance of kanamycin-resistant variants

Frequency of appearance of kanamycin-resistant variants was determined by subculturing strains PA14 or PA14 carrying pED202 in LB liquid medium, and incubating for 15–16 h at 37 °C. Cultures were plated on LB, LB without NaCl or M63 containing kanamycin (200 µg ml⁻¹) after plates were dried for at least 1 h. Plates were incubated 36–40 h at 37 °C or 25 °C before resistant variants were counted.

Analysis of sputum samples from CF patients

Sputum samples were suspended in 5 ml of 10 mM MgSO₄. Serial dilutions of samples were plated onto Cetrimide agar (Difco) with and without antibiotics. Plates were screened for the presence of *P. aeruginosa* after 48 h of incubation at 37 °C. Colonies were later verified to be *P. aeruginosa* by probing colony lifts with the exotoxin A gene from *P. aeruginosa* (contained in the EcoRI–HindIII fragment of plasmid pRGI (ref. 24)).

Antibiotic susceptibility of reverted variants

Variants obtained from CF samples were grown in 5 ml LB agar overnight at 37 °C. After incubation, serial dilutions of the bacterial suspension were plated onto antibiotic-free LB agar, and incubated overnight at 37 °C, and subsequently for 5 days at 25 °C. Plates were examined for the presence of revertants at the edges of the colonies. We repeated this process until stable revertants were obtained. *In vitro* susceptibility to the different antibiotics used was determined by establishing minimal bactericidal concentrations (MBCs) for the reverted variants as described²⁵.

Cloning of genetic region controlling phenotypic variation

A PA14 genomic library in vector pJSR1 (ref. 7) was mobilized en masse into RSCV PA14 using strain pRK2013 as helper²⁶. Cosmid pED20 was subcloned into the pUCP19 plasmid vector²⁷ using a *Pst*I restriction digest.

Construction of a *P. aeruginosa* pvrR mutant

A *pvrR* deletion was generated by replacing 2.33 kb of the wild-type sequence containing the *pvrR* gene with a 1.416-kb fragment amplified by PCR. The PCR-amplified fragment was subcloned into the *Xba*I and *Sma*I sites of pCVD442 (ref. 28), generating pED167,

which was introduced by allelic exchange into the homologous region of the PA14 chromosome as described²⁸.

Received 7 January; accepted 12 February 2002.

- Govan, J. R. & Deretic, V. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* **60**, 539–574 (1996).
- Smith, J. J., Travis, S. M., Greenberg, E. P. & Welsh, M. J. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* **85**, 229–236 (1996).
- Hoiby, N. Antibiotic therapy for chronic infection of pseudomonas in the lung. *Annu. Rev. Med.* **44**, 1–10 (1993).
- Singh, P. K. *et al.* Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764 (2000).
- Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
- Speert, D. P. *et al.* Conversion of *Pseudomonas aeruginosa* to the phenotype characteristic of strains from patients with cystic fibrosis. *J. Clin. Microbiol.* **28**, 188–194 (1990).
- Rahme, L. G. *et al.* Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268**, 1899–1902 (1995).
- Hausler, S., Tummeler, B., Weissbrodt, H., Rohde, M. & Steinmetz, I. Small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis. *Clin. Infect. Dis.* **29**, 621–625 (1999).
- Drumm, B., Neumann, A. W., Policova, Z. & Sherman, P. M. Bacterial cell surface hydrophobicity properties in the mediation of *in vitro* adhesion by the rabbit enteric pathogen *Escherichia coli* strain RDEC-1. *J. Clin. Invest.* **84**, 1588–1594 (1989).
- Hentzer, M. *et al.* Alginate overproduction affects *Pseudomonas aeruginosa* biofilm structure and function. *J. Bacteriol.* **183**, 5395–5401 (2001).
- Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches—the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33**, 919–932 (1999).
- Grewal, S. I., Han, B. & Johnstone, K. Identification and characterization of a locus which regulates multiple functions in *Pseudomonas tolaasii*, the cause of brown blotch disease of *Agaricus bisporus*. *J. Bacteriol.* **177**, 4658–4668 (1995).
- Holloway, B. W., Krishnapillai, V. & Morgan, A. F. Chromosomal genetics of *Pseudomonas*. *Microbiol. Rev.* **43**, 73–102 (1979).
- Oliver, A., Canton, R., Campo, P., Baquero, F. & Blazquez, J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**, 1251–1254 (2000).
- Han, B., Pain, A. & Johnstone, K. Spontaneous duplication of a 661 bp element within a two-component sensor regulator gene causes phenotypic switching in colonies of *Pseudomonas tolaasii*, cause of brown blotch disease of mushrooms. *Mol. Microbiol.* **25**, 211–218 (1997).
- Lewis, K. Riddle of biofilm resistance. *Antimicrob. Agents Chemother.* **45**, 999–1007 (2001).
- Mah, T. F. & O'Toole, G. A. Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol.* **9**, 34–39 (2001).
- McNamara, P. J. & Proctor, R. A. *Staphylococcus aureus* small colony variants, electron transport and persistent infections. *Int. J. Antimicrob. Agents* **14**, 117–122 (2000).
- Ausubel, F. M. *et al.* *Short Protocols in Molecular Biology* (Greene Publishing and Wiley, New York, 1992).
- O'Toole, G. A. & Kolter, R. Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Mol. Microbiol.* **28**, 449–461 (1998).
- Sherman, P. M., Houston, W. L. & Boedeker, E. C. Functional heterogeneity of intestinal *Escherichia coli* strains expressing type 1 somatic pili (fimbriae): assessment of bacterial adherence to intestinal membranes and surface hydrophobicity. *Infect. Immun.* **49**, 797–804 (1985).
- Christensen, B. B. *et al.* Molecular tools for study of biofilm physiology. *Methods Enzymol.* **310**, 20–42 (1999).
- Bloemberg, G. V., O'Toole, G. A., Lugtenberg, B. J. & Kolter, R. Green fluorescent protein as a marker for *Pseudomonas* spp. *Appl. Environ. Microbiol.* **63**, 4543–4551 (1997).
- Samadpour, M., Moseley, S. L. & Lory, S. Biotinylated DNA probes for exotoxin A and pilin genes in the differentiation of *Pseudomonas aeruginosa* strains. *J. Clin. Microbiol.* **26**, 2319–2323 (1988).
- Andrews, J. M. Determination of minimum inhibitory concentrations. *J. Antimicrob. Chemother.* **48**, (Suppl. A) 5–16 (2001).
- Figurski, D. H. & Helinski, D. R. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl Acad. Sci. USA* **76**, 1648–1652 (1979).
- Schweizer, H. P. *Escherichia-Pseudomonas* shuttle vectors derived from pUC18/19. *Gene* **97**, 109–121 (1991).
- Donnenberg, M. S. & Kaper, J. B. Construction of an eae deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect. Immun.* **59**, 4310–4317 (1991).

Acknowledgements

We thank J. Lai and R. A. Rogers for image analysis of biofilm structures; T. Mylonakis for help in obtaining the samples and clinical data from CF patients; M. J. Ferraro and K. Henry for providing the CF sputum samples; and M. Wildermuth for help analysing protein sequences. This work was supported by grants from Aventis SA and the National Institutes of Health.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to F.M.A. (e-mail: ausubel@molbio.mgh.harvard.edu). The sequences for *pvrR*, ORF1 and ORF3 are available at GenBank under accession number AF482691.

RANKL maintains bone homeostasis through c-Fos-dependent induction of interferon- β

Hiroshi Takayanagi*, Sunhwa Kim*, Koichi Matsuo†‡, Hiroshi Suzuki*, Tomohiko Suzuki§, Kojiro Sato*, Taeko Yokochi*, Hiromi Oda||, Kozo Nakamura||, Nobutaka Ida§, Erwin F. Wagner† & Tadatsugu Taniguchi*

* Department of Immunology, || Department of Orthopaedic Surgery, Faculty of Medicine and Graduate School of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

† Research Institute of Molecular Pathology (IMP), Dr. Bohr Gasse 7, A-1030 Vienna, Austria

‡ National Institute for Longevity Sciences (NILS), Obu, Aichi 474-8522, Japan

§ Pharmaceutical Research Laboratories, Toray Industries, Inc., Kamakura, Kanagawa 248-8555, Japan

Osteoclasts are cells of monocyte/macrophage origin that erode bone matrix: regulation of their differentiation is central to the understanding of the pathogenesis and treatment of bone diseases such as osteoporosis^{1,2}. Signalling by RANKL (receptor activator of NF- κ B ligand), also known as *Tnfrsf11*, is essential for the induction of osteoclast differentiation^{3–5}, and it must be strictly regulated to maintain bone homeostasis. But it is not known whether RANKL signalling to the cell interior is linked to any regulatory mechanisms. Here we show that RANKL induces the interferon- β (*IFN- β*) gene in osteoclast precursor cells, and that IFN- β inhibits the differentiation by interfering with the RANKL-induced expression of c-Fos, an essential transcription factor for the formation of osteoclasts. This *IFN- β* gene induction mechanism is distinct from that induced by virus, and is dependent on c-Fos itself. Thus an autoregulatory mechanism operates—the RANKL-induced c-Fos induces its own inhibitor. The importance of this regulatory mechanism for bone homeostasis is emphasized by the observation that mice deficient in IFN- β signalling exhibit severe osteopenia (loss of bone mass) accompanied by enhanced osteoclastogenesis. Our study places the IFN- β system in a new context, and may offer a molecular basis for the treatment of bone diseases.

Bone remodelling depends on a delicate balance between bone formation and bone resorption, wherein bone-forming osteoblasts and bone-resorbing osteoclasts play central roles¹. Indeed, tipping this balance in favour of osteoclasts leads to pathological bone resorption, as seen in bone diseases such as autoimmune arthritis and postmenopausal osteoporosis^{2,6}. RANKL, a member of the tumour necrosis factor (TNF) family, is essential for osteoclastogenesis, and its signalling mechanism has been extensively studied. Briefly, binding of RANKL to its receptor, RANK, results in the recruitment of TRAF family proteins such as TRAF6 (TNF receptor-associated factor 6), which activates the NF- κ B and JNK pathways^{5,7}. RANKL also induces the expression of c-Fos by an as yet unknown mechanism^{8,9}. The essential role of the TRAF6 and c-Fos pathways in osteoclastogenesis determined by gene disruption studies has been well documented^{10–13}. In addition, negative regulators of RANKL signalling have been reported^{7,14}. A typical example is osteoprotegerin (OPG), which functions as a soluble 'decoy' receptor for RANKL, thereby attenuating excessive RANKL signalling^{5,15,16}. It is, however, currently unknown whether RANKL signalling itself activates a negative regulatory mechanism to maintain the bone homeostasis.

During the course of our study to identify target genes induced by RANKL in osteoclast precursor cells, we found the induction of messenger RNAs, whose expression is dependent on IFN- α/β

signalling. The induction was notably decreased in cells lacking c-Fos; the downregulated genes include *Ifit3* (also known as *GARG49*, *IRG2*), *Ifit2* (*GARG39*), *Mx1*, *Scyb10* (*CRG2*, *IP10*) and *Irf7* (K.M. and H.T., unpublished data). Although the integral role of the IFN- α/β system in the regulation of immune responses has been extensively documented, it is unknown whether this system is linked to RANKL signalling and bone metabolism. To address this issue, we first evaluated the skeletal system of mice deficient in one of the IFN- α/β receptor components, *IFNAR1* (*IFNAR1*^{−/−} mice)¹⁷. A notable reduction in trabecular bone mass, a characteristic feature of osteoporosis, was observed in the mutant mice (Fig. 1a). Concomitantly, the number of multinucleated cells (MNCs) positive for the osteoclast marker, tartrate-resistant acid phosphatase (TRAP⁺ MNCs), is notably increased in the bones of *IFNAR1*^{−/−} mice compared to wild-type (WT) mice (Fig. 1b). Quantitative bone morphometric analyses further revealed a significant decrease in the trabecular bone volume, diagnostic for enhanced osteoclastic bone resorption, without any significant difference in the values of markers of osteoblastic bone formation (Fig. 1c). These results suggest that IFN- α/β signalling is physiologically critical for the maintenance of normal bone mass through negative regulation of osteoclastic bone resorption.

As shown in Fig. 1d, *in vitro* osteoclast formation induced by RANKL in the presence of macrophage colony-stimulating factor (M-CSF) is markedly enhanced in bone marrow monocyte/macrophage precursor cells (BMMs) from *IFNAR1*^{−/−} mice. On the other hand, cultured calvarial cells derived from WT and *IFNAR1*^{−/−} mice showed no difference in osteoblast differentiation (Fig. 1e) and in their ability to support osteoclastogenesis (data not shown). These *in vitro* results are consistent with the above *in vivo* observation, collectively suggesting that the enhanced osteoclastogenesis in *IFNAR1*^{−/−} mice is due to a cell-autonomous abnormality of BMMs. Although it cannot be rigorously excluded that IFN- β produced by other cell types, such as T cells, additionally contributes to the *in vivo* phenotype, there is currently no evidence that suggests this possibility (Supplementary Information Fig. 1 and H.T., unpublished data).

The above findings suggest that the IFN- α/β system is critical for the negative feedback regulation of osteoclastogenesis, and that RANKL may induce IFN- α/β gene expression in BMMs. RANKL stimulation of BMMs resulted in the induction of mRNA for IFN- β , but not for IFN- α (Fig. 2a) as detected by semiquantitative reverse transcription–polymerase chain reaction (RT–PCR). The results thus suggest that, unlike the common antiviral function of these two IFN subtypes, IFN- β is selectively involved in osteoclast regulation. Accordingly, we examined the skeletal system of mice lacking IFN- β (*IFN- β* ^{−/−} mice)¹⁸. Similar to *IFNAR1*^{−/−} mice, these mice also exhibit severe osteopenia resulting from enhanced osteoclastogenesis, indicating the critical role of RANKL-induced IFN- β in bone homeostasis (Fig. 2b).

Recombinant mouse IFN- β strongly inhibited osteoclastogenesis from BMMs stimulated by RANKL in the presence of M-CSF: its profound inhibitory effect was observed at a concentration as low as 1 U ml^{−1} (Fig. 2c). At this concentration, IFN- β had little, if any, effect on M-CSF-induced proliferation (data not shown). The same inhibitory effect of IFN- β on osteoclastogenesis was observed in RAW 264.7 cells, a macrophage cell line that differentiates into osteoclasts in response to RANKL (data not shown). These results suggest that IFN- β interferes with RANKL signalling, thereby inhibiting osteoclastogenesis. IFN- β (as well as IFN- α) elicits cellular responses through activation/induction of transcription factors, which include ISGF3 (a heterotrimeric complex consisting of Stat1, Stat2 and interferon regulatory factor 9 (IRF-9)), GAF/AAF (the Stat1 homodimer) and IRF-1 (refs 19, 20). The inhibitory action of IFN- β is abrogated in BMMs of mice lacking Stat1 or IRF-9, but not IRF-1 (Fig. 2d), indicating that the inhibitory action is linked to the ISGF3-mediated gene induction pathway.

To gain insights into the mechanism of the IFN- β inhibition of the RANKL-induced osteoclastogenesis, we carried out immunoblot analysis of molecules that are known to be critical effectors of RANKL signalling, in RANKL-stimulated BMMs with or without exogenous IFN- β treatment. The RANKL-induced expression of c-Fos, but not other molecules examined, was downregulated by IFN- β (Fig. 3a, and Supplementary Information Fig. 3a). Consistent with the normal expression of TRAF6 (Fig. 3a), the activation of its downstream molecules, NF- κ B, JNK and p38, occurred

normally even in the presence of IFN- β , as determined by electrophoretic mobility shift assay (EMSA) or immunoblot analysis⁷ (see Supplementary Information Figs 2 and 3b). These observations prompted us to examine whether ectopic expression of c-Fos can rescue BMMs from the IFN- β inhibition of osteoclast differentiation, using retrovirus-mediated gene transfer. As shown in Fig. 3b, c-Fos overexpression efficiently rescued the differentiation blockade by IFN- β . These results suggest that the down-regulation of c-Fos accounts, at least in part, for the IFN- β

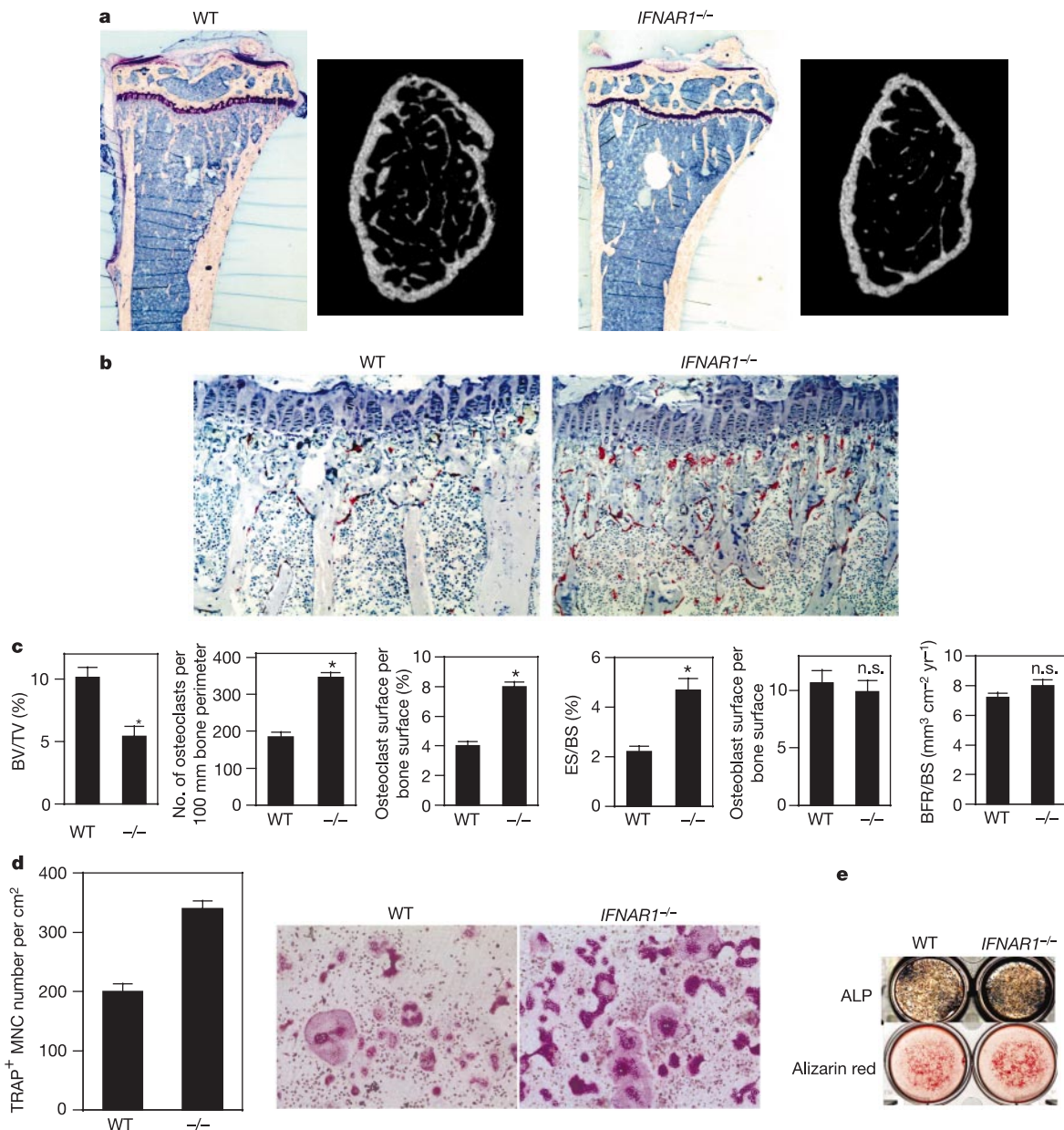


Figure 1 Bone phenotype of mice deficient in IFN- α/β signalling. **a**, Histological and radiological analyses of bone of wild-type (WT) and *IFNAR1*^{-/-} mice. Reduction of the trabecular bone volume is evident in a longitudinal section (toluidine-blue staining) of the tibia, and a transverse section (microcomputed tomography) of the metaphyseal region of the distal femur (10% of length above the growth plate) of *IFNAR1*^{-/-} mice at 12 weeks of age. The same degree of reduction of bone volume was also observed at 24 weeks of age. **b**, Increased number of TRAP⁺ (red-stained) multinucleated cells (MNCs) below the growth plate in *IFNAR1*^{-/-} mice. **c**, Histomorphometrical analysis of *IFNAR1*^{-/-} mice ($n = 5$, data shown as mean $\pm$ s.e. * $P < 0.01$) demonstrates a significant reduction of trabecular bone volume, BV/TV, accompanied by enhanced values of indicators of

osteoclastic bone resorption (number of osteoclasts per bone perimeter, osteoclast surface per bone surface, and erosive surface per bone surface (ES/BS)), but there was no difference in the markers of osteoblastic bone formation (osteoblast surface per bone surface, and bone formation rate per bone surface, BFR/BS). A reduction of cortical bone volume (nearly 20%) was also observed. **d**, *In vitro* differentiation of osteoclasts is cell-autonomously upregulated in *IFNAR1*^{-/-} bone marrow monocyte/macrophage precursor cells (BMMs). Note the increased number of TRAP⁺ MNCs and enhanced multinuclearity in *IFNAR1*^{-/-} cells. **e**, Normal osteoblast differentiation of *IFNAR1*^{-/-} calvarial cells examined by the alkaline phosphatase (ALP) and alizarin red staining. n.s., not significant.

inhibition of osteoclastogenesis.

What is the mechanism underlying the inhibition of c-Fos expression by IFN- β ? The RANKL-induced c-Fos mRNA was not significantly altered by IFN- β , suggesting a post-transcriptional control mechanism(s) (Fig. 3c). Consistent with previous reports^{8,9}, the Fra-1 and Fra-2 mRNA expression levels, which are controlled by c-Fos, were also downregulated (Fig. 3c). As expected, immunoblot analysis revealed that the expression of Fra-1 and Fra-2 proteins was severely downregulated (data not shown). To gain further insights into the mechanism of c-Fos suppression by IFN- β , the synthesis of c-Fos was examined by pulse-chase experiments. As shown in Fig. 3d, c-Fos protein synthesis was inhibited by IFN- β in RANKL-stimulated BMMs.

In view of the results showing that the IFN- β action is mediated by ISGF3 and by inhibition of c-Fos synthesis, one candidate mediator is the double-stranded-RNA-activated protein kinase (PKR). PKR, induced by IFN- β through ISGF3 activation, is

involved in the inhibition of protein synthesis by phosphorylation of eIF2 α , an initiation factor of translation²¹. We then examined the inhibitory effect of IFN- β on RANKL-induced osteoclastogenesis in mice lacking PKR (*PKR*^{-/-} mice)²² (Fig. 3e), and observed that the suppressive effect is attenuated, albeit only partially. Thus, PKR is likely to be one of the target genes of ISGF3 responsible for the IFN- β -mediated inhibition of osteoclastogenesis.

It has been shown that the induction of IFN- α and - β genes in virus-infected cells requires two IRF transcription factors, IRF-3 and IRF-7 (refs 19, 23, 24). In the case of IFN- β , NF- κ B is also involved in this induction²³. However, as mentioned above, the induction of IFN-inducible genes was not observed in osteoclast precursor cells lacking c-Fos (K.M., unpublished data), suggesting a hitherto unrecognized function of c-Fos in the RANKL induction of the IFN- β gene. To address this issue, we examined the RANKL-induced IFN- β mRNA expression in osteoclast precursor cells from mice lacking c-Fos (*Fos*^{-/-} mice)¹², or both IRF-3 and IRF-9 (*IRF*-

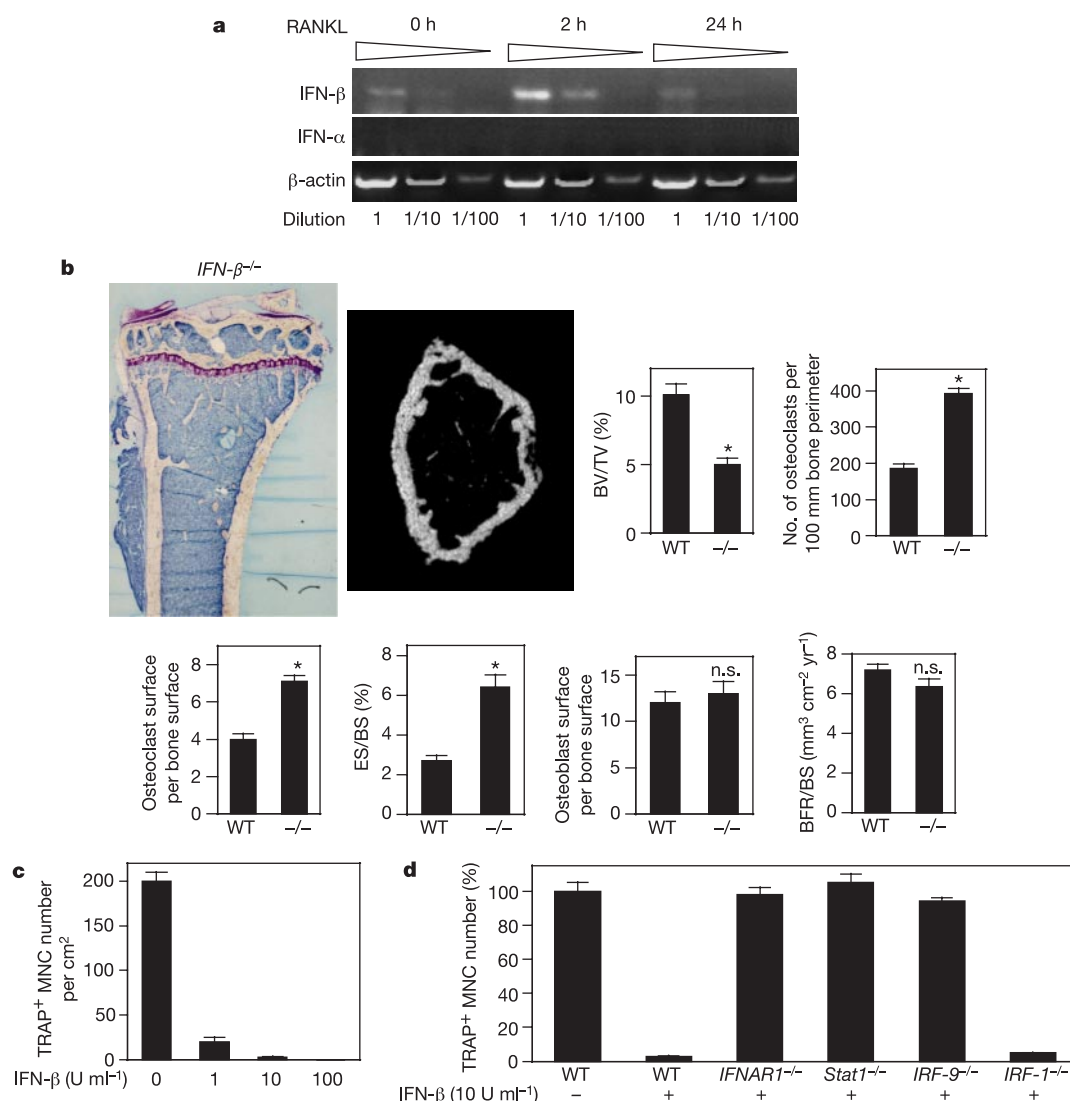


Figure 2 Selective induction of IFN- β mRNA by RANKL and the function of IFN- β *in vivo* and *in vitro*. **a**, Semiquantitative RT-PCR analysis of IFN- α and IFN- β mRNAs in BMMs after RANKL stimulation. The dilution rate is indicated below each lane. The IFN- β mRNA induction level was estimated to be more than 20-fold by the quantitative real-time PCR system. This induction is weaker (about one order) than that of virus-infected cells. **b**, Reduction of trabecular bone volume and increased number of osteoclasts in *IFN- β* ^{-/-} mice, examined as described in Fig. 1 (* *P* < 0.01). No abnormality was found in the

osteoblastic bone formation in *IFN- β* ^{-/-} mice. **c**, Inhibitory effect of IFN- β on osteoclastogenesis from BMMs stimulated by RANKL with M-CSF. An inhibitory effect of IFN- α on osteoclastogenesis was less pronounced: approximately 10-fold higher concentration was required to achieve the same level of the inhibition (data not shown). **d**, Effect of IFN- β on RANKL-induced osteoclastogenesis in BMMs derived from mice deficient in either IFNAR1, Stat1, IRF-9 or IRF-1.

3/IRF-9^{-/-} mice; the IRF-9 deficiency leads to the loss of IRF-7 expression²⁴). Splenocytes from IRF-3/IRF-9^{-/-} mice were totally defective in IFN-β gene induction by viruses²⁴, and the same observation was made in BMMs (data not shown). The induction of IFN-β mRNA by RANKL was still detectable in osteoclast

precursor cells from IRF-3/IRF-9^{-/-} mice, but it was no longer observed in the cells from Fos^{-/-} mice (Fig. 4a). On the other hand, virus-induced expression of IFN-β mRNA was unaffected in splenocytes from Fos^{-/-} mice (data not shown).

We also examined the activation of the IFN-β promoter by

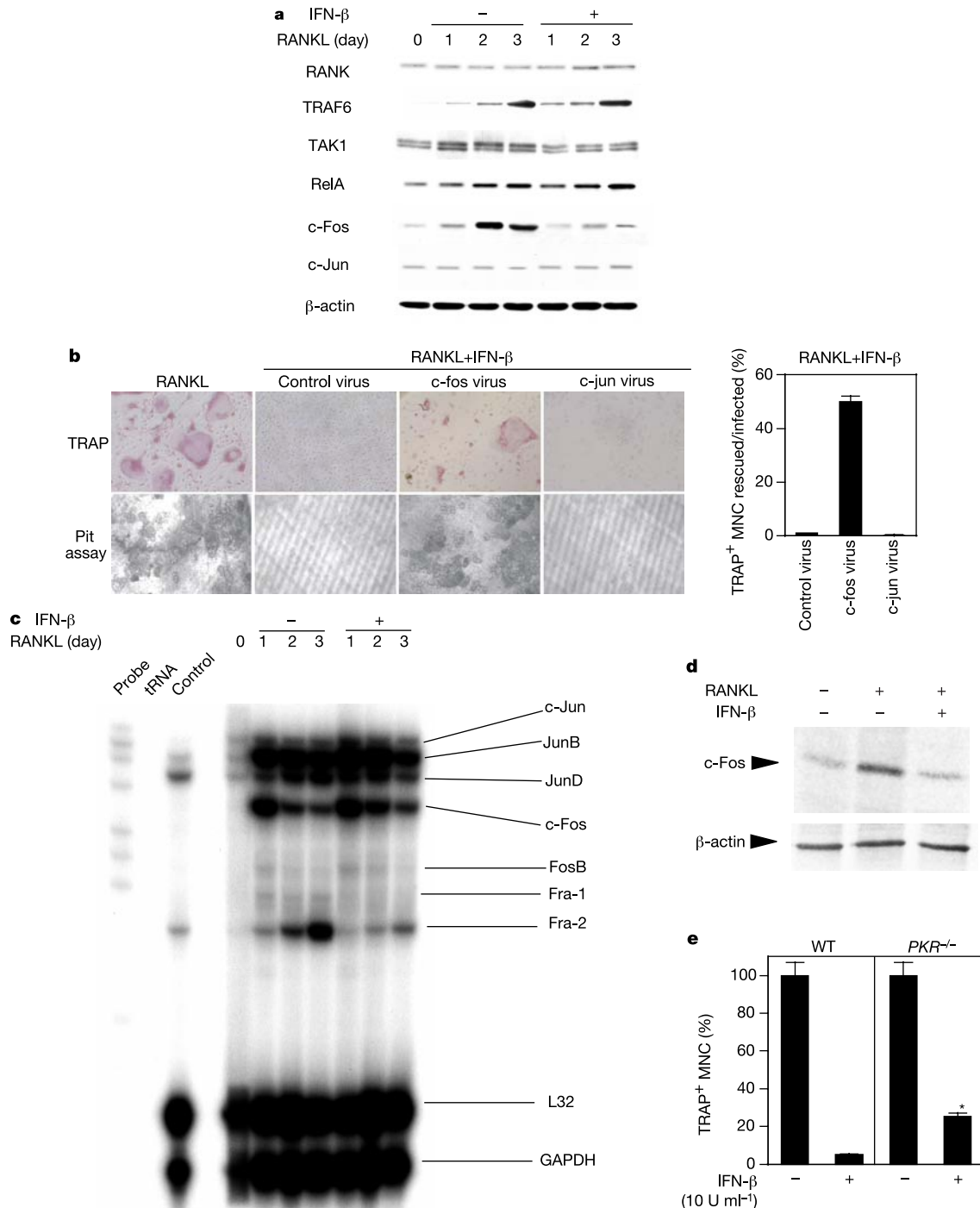


Figure 3 IFN-β inhibits RANKL signalling by suppressing the c-Fos protein expression. **a**, Immunoblot analysis of downstream effector molecules involved in RANKL signalling. IFN-β had no effect on the expression of other NF-κB components such as p50 and p52 (data not shown). **b**, Attenuation of the IFN-β-mediated inhibition of osteoclastogenesis by ectopic expression of c-Fos. Decreased formation of TRAP⁺ MNCs and resorption pits is attenuated by infecting the c-fos virus (pBabe-c-fos), but not by control (pBabe-EGFP) or c-jun (pBabe-c-jun) viruses⁸. More than 50% of the cells infected with the c-fos virus became TRAP-positive in the presence of IFN-β (10 U ml⁻¹), under which condition less than 5% of the cells infected by either control or c-jun virus became TRAP-positive.

Consistent with a previous report⁸, infection with retroviruses expressing Fra-1 or Fra-2 also caused resistance to IFN-β-mediated inhibition of osteoclastogenesis (data not shown). **c**, Effect of IFN-β on the mRNA expression of AP-1 components (RNase protection assay). There was no significant difference in the expression of Jun family members. Downregulation of Fra-1 and Fra-2 mRNA was further confirmed by semiquantitative RT-PCR (data not shown). **d**, Inhibition of c-Fos protein synthesis by IFN-β (pulse chase assay). **e**, Effect of IFN-β on osteoclastogenesis from BMMs derived from PKR^{-/-} mice. (* *P* < 0.05 versus WT).

RANKL by transfecting the promoter-driven luciferase reporter gene (*p125Mu-Luc*)²⁴ into RAW 264.7 cells. As shown in Fig. 4b, RANKL stimulation of the transfected cells resulted in a significant induction of luciferase activity. A similar induction was also observed by ectopic expression of c-Fos instead of RANKL stimulation (Fig. 4b). Furthermore, this induction was abolished when the putative c-Fos binding site of this promoter (AP-1 site), which has been shown to bind c-Jun/ATF-2 in virus-infected cells²³, was mutated (Fig. 4b). Using the AP-1 site in the *IFN-β* promoter as a probe, a specific DNA binding activity was detected by EMSA in extracts from RANKL-stimulated BMMs, and this activity was inhibited by an anti-c-Fos antibody (see Supplementary Information Fig. 4). Thus, c-Fos may be directly involved in the activation of the *IFN-β* promoter. In addition, the RANKL-activated NF-κB, the binding site of which is found in the promoter of *IFN-β*, but not *IFN-α*, may cooperatively contribute to this activation. Taking these observations together, RANKL induction of *IFN-β* must be considered distinct in terms of both induction mechanism and biological function.

These observations prompted us to examine the *in vivo* efficacy of *IFN-β* for the treatment of osteoclast-mediated pathological conditions, using a model of endotoxin-induced inflammatory bone destruction⁷. Administration of *IFN-β* into the site of inflammation resulted in marked inhibition of osteoclast formation and bone resorption (Fig. 5a). Thus, exogenous application of *IFN-β* indeed has a beneficial effect against bone destruction in this disease model, although *in vivo* action of *IFN-β* may extend beyond the inhibition of RANKL signalling. In addition, our preliminary results also

suggest that systemic administration of *IFN-β* can reverse the bone loss in an osteoporosis model of ovariectomized mice (T.S., N.I. and H.T., unpublished observation).

This study revealed an unexpected aspect of *IFN-β* induction and function, beyond its original context of host defence against virus and other pathogens^{19,20}. In fact, a hitherto unknown signalling cross-talk between RANKL and *IFN-β* system via c-Fos operates to maintain bone homeostasis. This cross-talk is unique in that RANKL signalling *per se* is responsible for the *IFN-β* gene induction, and that the positive mediator of RANKL signalling, c-Fos, is required for the induction of its own inhibitor (Fig. 5b). It is known that osteoclastogenesis is also under negative regulation by OPG, the loss of which similarly causes an osteoporotic phenotype in mice¹⁶. The negative regulation by *IFN-β* is distinct from that by OPG in that the osteoclast precursor cells themselves produce the negative regulator in response to the positive signalling, and that *IFN-β* interferes with the intracellular signalling pathway of RANKL. Thus, both regulatory mechanisms, functioning at different levels, are important to maintain the homeostatic bone

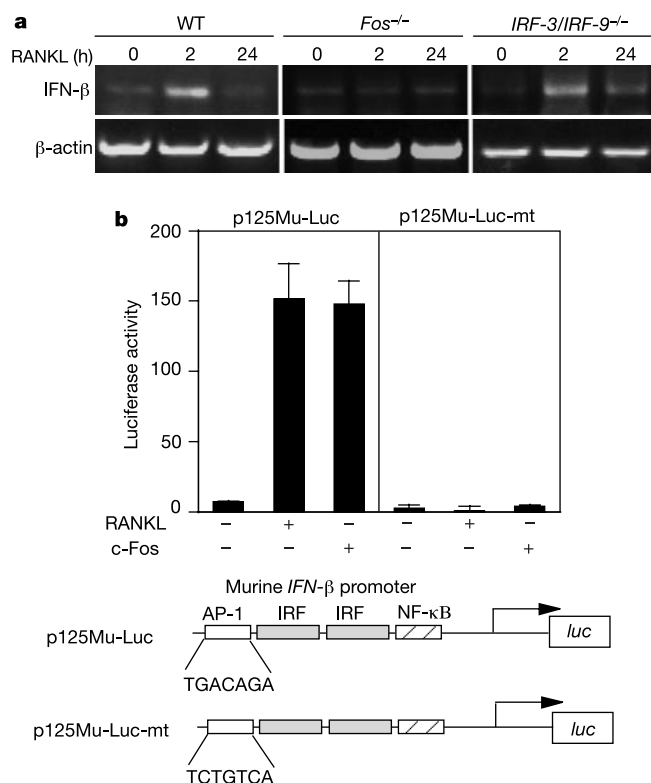
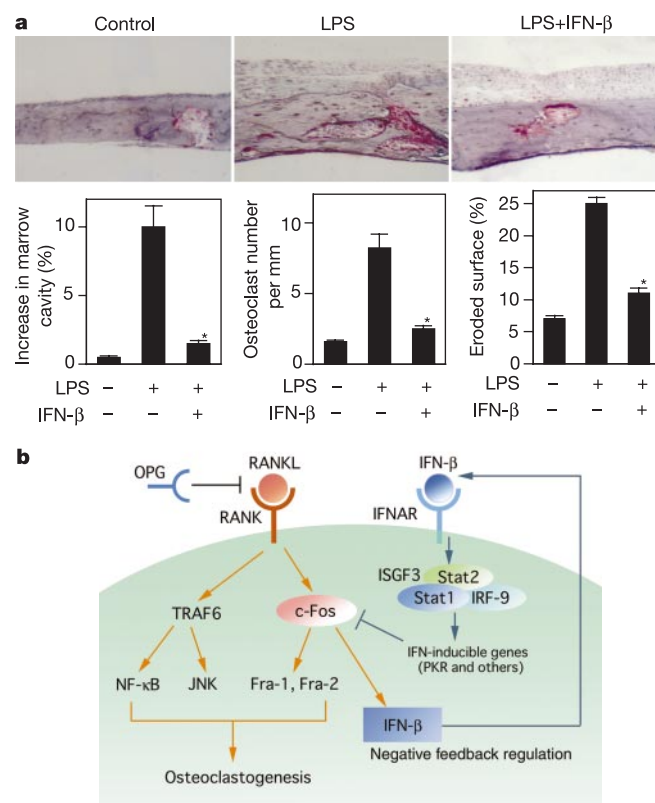


Figure 4 Fos-dependent induction of *IFN-β* by RANKL. **a**, Induction of *IFN-β* mRNA by RANKL (RT-PCR analysis). The induction is observed in BMMs (and splenocytes; data not shown) from *IRF-3/IRF-9*^{-/-} mice, but not observed in osteoclast precursor cells from *Fos*^{-/-} mice. **b**, Activation of *IFN-β* promoter by RANKL and c-Fos. The induction of luciferase activity is shown in the upper panel. Location of the binding sites for AP-1, IRFs and NF-κB within the *IFN-β* promoter, and the sequence of the AP-1 site is depicted in the lower panel.



resorption. The present work may also offer approaches to the therapy of bone diseases caused by excessive osteoclastogenesis, such as autoimmune arthritis and osteoporosis^{1,2,25}. □

Methods

Mice and bone phenotype analysis

We purchased *IFNAR1*^{-/-} mice¹⁷ from the B&K Universal Group, and generated *IFN-β*^{-/-} mice as described¹⁸. These mutant mice were backcrossed with C57BL/6 mice more than 10 times, and subjected to histomorphometric and microradiographic examinations as described²⁶. They showed no abnormality in growth rate or body weight. The bone phenotypes of WT littermates of both mutants are essentially the same as those of inbred C57BL/6 mice, and they served as WT controls. We generated *IRF-1*^{-/-}, *IRF-9*^{-/-} and *Fos*^{-/-} mice as described^{12,27,28}. *Stat1*^{-/-} mice²⁹ and *PKR*^{-/-} mice²² were provided by R. D. Schreiber and C. Weissmann, respectively. All mice were born and kept under pathogen-free conditions. Statistical analysis was performed by Student's *t*-test.

In vitro differentiation of osteoclasts and osteoblasts

The method used for *in vitro* osteoclastogenesis was described previously⁷. Briefly, nonadherent bone marrow cells (5×10^5 per well in a 24-well plate) were cultured with 10 ng ml^{-1} M-CSF (R&D) for 2 d, and adherent cells were used as BMMs after washing out the nonadherent cells including lymphocytes. Monocyte/macrophage progenitor cells of *Fos*^{-/-} mice were derived from the spleen, and cultured as described⁶. These osteoclast precursor cells were further cultured in the presence of 100 ng ml^{-1} soluble RANKL (Peprotech) and 10 ng ml^{-1} M-CSF to generate osteoclasts; we invariably used these reagents at these concentrations. Recombinant mouse IFN-β ($3.5 \times 10^7 \text{ U mg}^{-1}$, Toray Inc.) and RANKL were added at the same time. Three days later, TRAP⁺ multinucleated (>3 nuclei) cells were counted. All data are expressed as mean $\pm$ s.e. ($n = 6$). TRAP⁺ MNCs were characterized by examining the bone-resorbing activity on dentine slices and the expression of calcitonin receptors as described³. *Ex vivo* osteoblast cultures were performed as described²⁶.

RT-PCR analysis of IFN-β mRNA

After a 2-d culture with M-CSF, BMMs were stimulated with RANKL for 2 or 24 h. RNA extraction and semiquantitative RT-PCR were performed as described³⁰, and the PCR product was sequenced to confirm the amplification of the *IFN-β* cDNA. Quantitative RT-PCR analysis was performed using a LightCycler and SyberGreen system (Roche).

Immunoblotting and RNase protection assay

To evaluate the expression of downstream molecules of RANKL signalling during osteoclastogenesis, whole-cell lysate of BMMs was assayed every 24 h after RANKL addition in the presence or absence of IFN-β (100 U ml^{-1}) by immunoblotting using specific antibodies for RANK, TRAF6, TAK1, RelA, p50, p52, c-Jun (Santa Cruz) and c-Fos (Calbiochem). In the RNase protection assay for the mRNAs of the AP-1 members, total RNA was extracted from BMMs every 24 h after RANKL addition and analysed using an RNase protection assay kit (Pharmingen) according to the manufacturer's protocol.

Pulse-chase labelling

After preincubation for 48 h in the presence of RANKL with or without IFN-β (100 U ml^{-1}), BMMs were pulsed for 30 min in the medium containing $100 \mu\text{Ci ml}^{-1}$ [³⁵S] methionine as described⁷. We carried out immunoprecipitation with a polyclonal antibody against c-Fos (Calbiochem) or β-actin (Sigma), followed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography.

Retroviral gene transduction

Construction of retrovirus vectors, pBabe-c-fos and pBabe-c-jun, has been described⁸. Retrovirus packaging was performed as described⁷. Two days after inoculation, BMMs were stimulated with RANKL and IFN-β (10 U ml^{-1}). Osteoclastogenesis was evaluated 3 d after RANKL addition by TRAP staining and bone resorption assay. We constructed pMX-c-fos-IRES-EGFP by inserting the *Bam*H I/*Eco*R I fragment of the mouse *Fos* gene derived from pBabe-c-fos. The rescue activity of TRAP⁺ cell formation in the virally infected cell population was determined by counting the TRAP⁺ cells within cells expressing the enhanced green fluorescent protein (EGFP) as described⁷.

Luciferase assay

The reporter plasmid, p125Mu-Luc, and the transfection method have been described²⁴, and the mutant p125Mu-Luc-mt was constructed by mutating the putative c-Fos binding site of the *IFN-β* gene promoter (-95 TGACAGA -89 to TCTGTCA). We stimulated RAW264.7 cells with RANKL 42 h after transfection. Luciferase assay was performed 48 h after transfection. All data are expressed as mean $\pm$ s.e. ($n = 4$).

Treatment of inflammatory bone destruction by IFN-β

Mice, 7–8 weeks old ($n = 20$), were administered a local calvarial injection of lipopolysaccharide (LPS) (Sigma) at 25 mg per kg weight. At the same site of LPS injection, the mice were subcutaneously administered with recombinant mouse IFN-β ($10,000 \text{ U d}^{-1}$) ($n = 10$) or control normal saline ($n = 10$) for 4 successive days. Five days after LPS injection, the number of osteoclasts per millimetre of trabecular bone surface, and the percentage of bone surface covered by osteoclasts (eroded surface) were measured as described⁷.

Received 27 November 2001; accepted 21 January 2002.

- Manolagas, S. C. Birth and death of bone cells: basic regulatory mechanisms and implications for the pathogenesis and treatment of osteoporosis. *Endocr. Rev.* **21**, 115–137 (2000).
- Rodan, G. A. & Martin, T. J. Therapeutic approaches to bone diseases. *Science* **289**, 1508–1514 (2000).
- Yasuda, H. *et al.* Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL. *Proc. Natl Acad. Sci. USA* **95**, 3597–3660 (1998).
- Kong, Y. Y. *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
- Suda, T. *et al.* Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr. Rev.* **20**, 345–357 (1999).
- Takayanagi, H. *et al.* Involvement of receptor activator of nuclear factor κB ligand/osteoclast differentiation factor in osteoclastogenesis from synovialocytes in rheumatoid arthritis. *Arthritis Rheum.* **43**, 259–269 (2000).
- Takayanagi, H. *et al.* T-cell-mediated regulation of osteoclastogenesis by signalling cross-talk between RANKL and IFN-γ. *Nature* **408**, 600–605 (2000).
- Matsuo, K. *et al.* *Fos1* is a transcriptional target of c-Fos during osteoclast differentiation. *Nature Genet.* **24**, 184–187 (2000).
- Wagner, E. F. & Karsenty, G. Genetic control of skeletal development. *Curr. Opin. Genet. Dev.* **11**, 527–532 (2001).
- Lomaga, M. A. *et al.* TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev.* **13**, 1015–1024 (1999).
- Naito, A. *et al.* Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice. *Genes Cells* **4**, 353–362 (1999).
- Wang, Z. Q. *et al.* Bone and haematopoietic defects in mice lacking c-fos. *Nature* **360**, 741–745 (1992).
- Grigoriadis, A. E. *et al.* c-Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**, 443–448 (1994).
- Roodman, G. D. Cell biology of the osteoclast. *Exp. Hematol.* **27**, 1229–1241 (1999).
- Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309–319 (1997).
- Bucay, N. *et al.* Osteoprotegerin-deficient mice develop early onset osteoporosis and arterial calcification. *Genes Dev.* **12**, 1260–1268 (1998).
- Müller, U. *et al.* Functional role of type I and type II interferons in antiviral defense. *Science* **264**, 1918–1921 (1994).
- Takaoka, A. *et al.* Cross talk between interferon-γ and -α/β signaling components in caveolar membrane domains. *Science* **288**, 2357–2360 (2000).
- Taniguchi, T., Ogasawara, K., Takaoka, A. & Tanaka, N. IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.* **19**, 623–655 (2001).
- Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–264 (1998).
- Samuel, C. E. The eIF-2 α protein kinases, regulators of translation in eukaryotes from yeasts to humans. *J. Biol. Chem.* **268**, 7603–7606 (1993).
- Yang, Y. L. *et al.* Deficient signaling in mice devoid of double-stranded RNA-dependent protein kinase. *EMBO J.* **14**, 6095–6106 (1995).
- Wathelet, M. G. *et al.* Virus infection induces the assembly of coordinately activated transcription factors on the IFN-β enhancer in vivo. *Mol. Cell* **1**, 507–518 (1998).
- Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN-α/β gene induction. *Immunity* **13**, 539–548 (2000).
- Takayanagi, H. *et al.* Suppression of arthritic bone destruction by adenovirus-mediated csk gene transfer to synovialocytes and osteoclasts. *J. Clin. Invest.* **104**, 137–146 (1999).
- Ogata, N. *et al.* Insulin receptor substrate-1 in osteoblast is indispensable for maintaining bone turnover. *J. Clin. Invest.* **105**, 935–943 (2000).
- Matsuyama, T. *et al.* Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development. *Cell* **75**, 83–97 (1993).
- Kimura, T. *et al.* Essential and non-redundant roles of p48 (ISGF3γ) and IRF-1 in both type I and type II interferon responses, as revealed by gene targeting studies. *Genes Cells* **1**, 115–124 (1996).
- Meraz, M. A. *et al.* Targeted disruption of the Stat1 gene in mice reveals unexpected physiologic specificity in the JAK-STAT signaling pathway. *Cell* **84**, 431–442 (1996).
- Hida, S. *et al.* CD8⁺ T cell-mediated skin disease in mice lacking IRF-2, the transcriptional attenuator of interferon-α/β signaling. *Immunity* **13**, 643–655 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank H. Murayama and Kureha Chemical Industries for technical assistance, and E. Barsoumian, A. Bichl, M. Radolf, A. Takaoka, K. Honda, M. Asagiri, N. Hata, S. Muraki, M. Isobe, S. Kano and I. Kawai for discussion and assistance. This work was supported in part by a grant for Advanced Research on Cancer from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, a research fellowship of the Japan Society for the Promotion of Science for Young Scientists, Health Sciences Research Grants from the Ministry of Health and Welfare of Japan, the Japan Orthopaedics and Traumatology Foundation, and PRESTO, Japan Science and Technology Corporation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (e-mail: tada@m.u-tokyo.ac.jp).

Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4

Nobutaka Suzuki[†], Shinobu Suzuki[†], Gordon S. Duncan^{*}, Douglas G. Millar[‡], Teiji Wada^{*}, Christine Mirtsos^{*}, Hidetoshi Takada^{*}, Andrew Wakeham^{*}, Annick Itie^{*}, Shyun Li[§], Josef M. Penninger^{*}, Holger Wesche[§], Pamela S. Ohashi[‡], Tak W. Mak^{*} & Wen-Chen Yeh^{*}

^{*}Amgen Institute, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 620 University Avenue, Suite 706, Toronto, Ontario M5G 2C1, Canada

[‡]Departments of Medical Biophysics and Immunology, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

[§]Tularik Incorporation, 2 Corporate Drive, South San Francisco, California 94080, USA

[†]These authors contributed equally to this work

Toll-like receptors (TLRs), which recognize pathogen-associated molecular patterns, and members of the pro-inflammatory interleukin-1 receptor (IL-1R) family, share homologies in their cytoplasmic domains called Toll/IL-1R/plant R gene homology (TIR) domains^{1–3}. Intracellular signalling mechanisms mediated by TIRs are similar⁴, with MyD88 (refs 5–8) and TRAF6 (refs 9, 10) having critical roles. Signal transduction between MyD88 and TRAF6 is known to involve the serine-threonine kinase IL-1 receptor-associated kinase 1 (IRAK-1)¹¹ and two homologous proteins, IRAK-2 (ref. 12) and IRAK-M¹³. However, the physiological functions of the IRAK molecules remain unclear, and gene-targeting studies have shown that IRAK-1 is only partially required for IL-1R and TLR signalling^{14,15}. Here we show by gene-targeting that IRAK-4, an IRAK molecule closely related to the *Drosophila* Pelle protein¹⁶, is indispensable for the responses of animals and cultured cells to IL-1 and ligands that stimulate various TLRs. IRAK-4-deficient animals are completely resistant to a lethal dose of lipopolysaccharide (LPS). In addition, animals lacking IRAK-4 are severely impaired in their responses to viral and bacterial challenges. Our results indicate that IRAK-4 has an essential role in innate immunity.

The signalling pathway downstream of mammalian Toll-like receptors is evolutionarily conserved with that initiated by the *Drosophila* Toll protein⁴. *Drosophila* Toll determines dorsoventral polarity in embryogenesis, and has a role in an anti-fungal immune response. In particular, transduction of intracellular signals initiated by the *Drosophila* Toll receptor system requires the serine-threonine kinase Pelle^{16,17}. In the mammalian genome, significant homology to Pelle and a parallel domain structure occurs in members of the IRAK family of intracellular signal transducers (Fig. 1a). These proteins are therefore promising candidates for an important role in mammalian TLR signalling.

IRAK-4 is the newest addition to the IRAK family (GenBank accession number AF445802)¹⁸. Among IRAK family members, IRAK-4 is most similar to Pelle in length, and the two proteins share 22% sequence identity. To characterize the physiological functions of IRAK-4, we generated *Irak-4*^{−/−} mice by gene targeting. A clone containing the genomic mouse *Irak-4* gene was isolated from a 129/J library and a targeting vector was constructed by replacing exon 2 (containing the start codon) with a neomycin-resistance cassette (Fig. 1b). Two independent and correctly targeted embryonic stem cell clones were used to generate *Irak-4*^{+/−} mice of a mixed 129/Ola; B6 background. These heterozygotes were intercrossed to produce *Irak-4*^{−/−} progeny, as confirmed by Southern blot analysis (Fig. 1c). Homozygous mutant mice were born at the expected mendelian ratio, and mice derived from both embryonic

stem cell lines had identical phenotypes. To facilitate analyses of the signalling cascade, we derived mouse embryonic fibroblasts (MEFs) from wild-type and homozygous mutant animals. Polymerase chain reaction with reverse transcription (RT-PCR) analysis using RNA collected from MEFs confirmed that *Irak-4* RNA expression is defective in *Irak-4*^{−/−} mice, although some residual aberrant RNA is detectable when the primer pair corresponding to the carboxy terminus of IRAK-4 is used (Fig. 1d). Western blot analysis using a specific antibody against mouse IRAK-4 demonstrated that neither intact nor truncated IRAK-4 proteins were present in *Irak-4*^{−/−} MEFs (Fig. 1e), confirming the null nature of the mutation.

IL-1 is a principal pro-inflammatory cytokine that is critical for innate immunity and which is highly induced after TLR signalling. To determine whether IRAK-4 has a role in IL-1 signalling, we examined several parameters associated with inflammatory responses. When IL-1β (40 μg kg^{−1} body weight) was injected intraperitoneally into wild-type mice, rapid, vigorous induction of both IL-6 (2,400 pg ml^{−1}) and tumour-necrosis factor-α (TNF-α, 135 pg ml^{−1}) was detected in serum after 2 h. However, this response was completely abolished in *Irak-4*^{−/−} mice (Fig. 2a). To determine whether this defect in the IL-1 response could be detected at the cellular level, we stimulated thioglycollate-elicited macrophages isolated from wild-type and mutant mice with IL-1 plus interferon-γ (IFN-γ). Activated macrophages from IRAK-4-deficient animals, in contrast to wild-type macrophages, were incapable of responding to IL-1 by producing nitric oxide (Fig. 2b). To rule out the possibility that this difference in the IL-

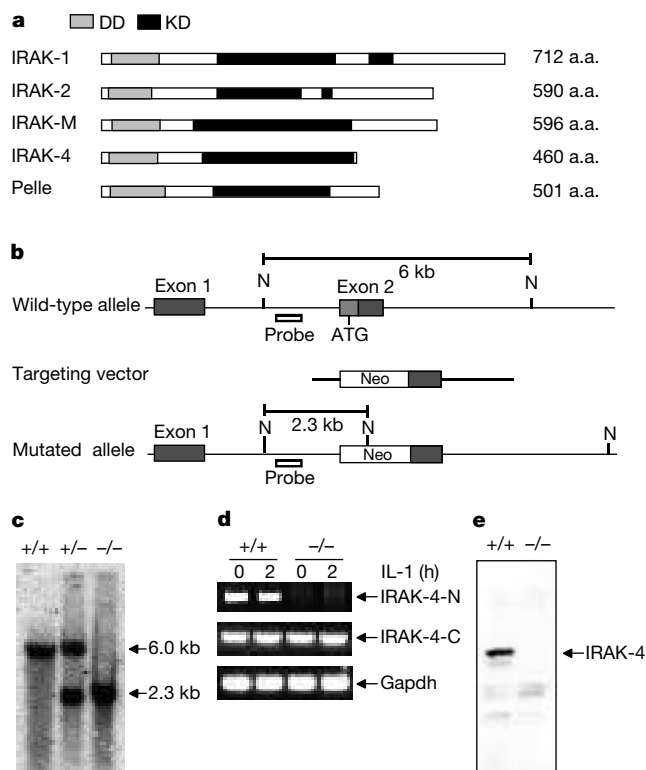


Figure 1 Generation of *Irak-4*^{−/−} mice. **a**, Domains of human IRAK proteins compared with *Drosophila* Pelle. DD, death domain; KD, kinase domain. Domains were determined using the Pfam 6.6 program. Length in amino acids is indicated. **b**, Partial structure of the mouse *Irak-4* gene, the targeting vector and the disrupted allele. The 5' flanking probe is shown. N, NcoI. **c**, Southern blot analysis of *Irak-4*^{+/−} intercross offspring. **d**, RT-PCR analysis of *Irak-4* RNA in IL-1-stimulated wild-type and *Irak-4*^{−/−} MEFs. PCR primers for the amino-terminus of the *Irak-4* gene (*Irak-4*-N) were derived from exon 2 and a downstream exon, and the primers for the *Irak-4*-C gene were derived from the carboxy-terminus. **e**, Western blot of IRAK-4 in wild-type and knockout MEFs.

1 response was due to a different level of activation in wild-type and *Irak-4*^{-/-} macrophages, we treated wild-type and *Irak-4*^{-/-} MEFs *in vitro* with IL-1. Again, severe defects in IL-6 and TNF- α production were observed in the mutant cells (Fig. 2c, left). In contrast, TNF- α treatment of wild-type or *Irak-4*^{-/-} MEFs resulted in comparable levels of IL-6 production (Fig. 2c, right). These results suggest that IRAK-4 is essential and specific for transducing cellular and animal responses to IL-1.

To demonstrate that the severe signalling defects observed in *Irak-4*^{-/-} cells were in fact due to the absence of wild-type IRAK-4, we stably transfected *Irak-4*^{-/-} MEFs with a murine full-length IRAK-4 expression vector and determined whether the mutant phenotype could be rescued. Indeed, *Irak-4*^{-/-} cells reconstituted with exogenous IRAK-4 were able to respond to IL-1 stimulation with a normal level of IL-6 induction, whereas wild-type MEFs stably over-expressing IRAK-4 maintained a normal (and not enhanced) IL-1 response (Fig. 2d).

Individual TLRs are responsible for recognizing specific microbial pathogen-associated molecular patterns (PAMPs), and constitute a chief molecular mechanism by which the host can respond to various pathogens. For example, TLR2, TLR3, TLR4, TLR5 and TLR9 receive and transmit signals in response to lipoproteins/peptidoglycans^{19,20}, viral RNA²¹, LPS^{22,23}, bacterial flagellin²⁴, and bacterial DNA²⁵, respectively. Among these ligands, LPS is the most studied because it induces both local and systemic inflammation, and responses to LPS mimic many of the features of tissue injury observed during infections with Gram-negative bacteria. Importantly, when injected into animals in large amounts, LPS triggers a mass production of pro-inflammatory cytokines that leads to clinical signs indistinguishable from septic shock.

As TLR signalling is highly similar to IL-1R signalling, we first examined whether IRAK-4 is involved in TLR4 signalling and in mediating LPS-induced septic shock. All wild-type mice injected

intraperitoneally with a lethal dose of LPS (40 mg kg⁻¹ body weight) died within 24 h of challenge (Fig. 3a). High concentrations of IL-6 (29 ng ml⁻¹), TNF- α (1.8 ng ml⁻¹) and IL-1 (6 ng ml⁻¹) were detected in sera of wild-type animals 2 h after LPS injection (Fig. 3b). In contrast, mice lacking IRAK-4 were completely resistant to LPS-induced septic shock and showed virtually no cytokine response (Fig. 3a, b). LPS treatment of purified B lymphocytes normally induces these cells to enter the cell cycle. Wild-type B cells treated with increasing concentrations of LPS showed increasing proliferative capacity, but *Irak-4*^{-/-} B cells were completely defective in this response (Fig. 3c). Furthermore, IRAK-4-deficient bone-marrow-derived macrophages (Fig. 3d) and MEFs (Fig. 3e) were also unresponsive to LPS stimulation as assessed by the production of various cytokines.

We next examined the effect of IRAK-4 deficiency on signalling through other TLRs by stimulating various cell types with the appropriate ligands. To determine the effect of IRAK-4 deficiency on TLR9 signalling, wild-type and *Irak-4*^{-/-} mice were injected intraperitoneally with oligodeoxynucleotides that contain unmethylated CpG motifs (CpG-ODN; 20 nmol) and D-galactosamine (20 mg). Similar to LPS, bacterial DNA was unable to stimulate IL-6 production in IRAK-4-deficient animals (Fig. 4a). Similarly, treatment of purified *Irak-4*^{-/-} B cells with CpG-ODN did not result in a proliferative response (Fig. 4b). These results indicate that TLR9 signalling requires IRAK-4 function. Furthermore, IRAK-4-deficient macrophages were unable to produce inflammatory mediators in response to treatment with either peptidoglycan (PGN) or poly(I:C) (Fig. 4c), indicating that TLR2

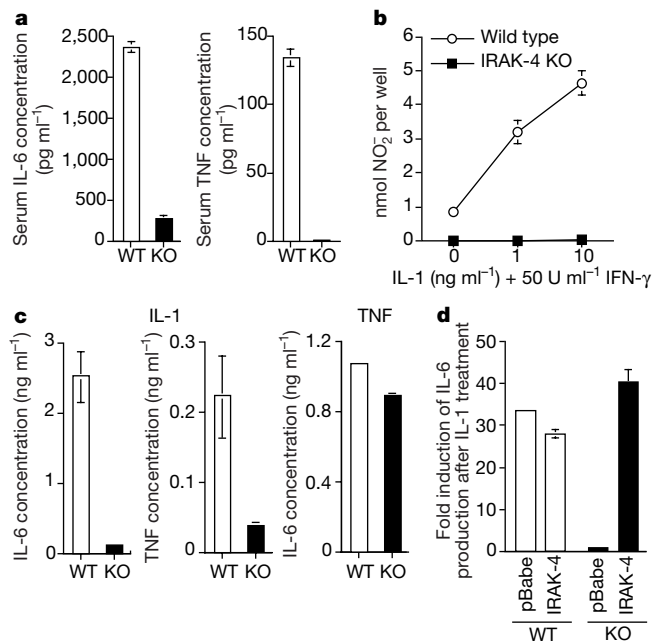


Figure 2 Impaired responses to IL-1 in *Irak-4*^{-/-} mice and cells. **a**, Serum levels of IL-6 and TNF- α in wild-type (WT) and knockout (KO) mice treated with IL-1. For all figures, results shown are the mean $\pm$ s.d. of triplicate determinations. **b**, Nitrite production in culture supernatants of IL-1-stimulated macrophages. **c**, Response of *Irak-4*^{-/-} MEFs to stimulation with 10 ng ml⁻¹ IL-1 or TNF- α . **d**, Rescue of the *Irak-4*^{-/-} phenotype by exogenous IRAK-4 expression. pBabe, vector-only control; IRAK-4, vector expressing full-length mouse IRAK-4. Results are shown as the number of fold increase in IL-6 induction over untreated controls.

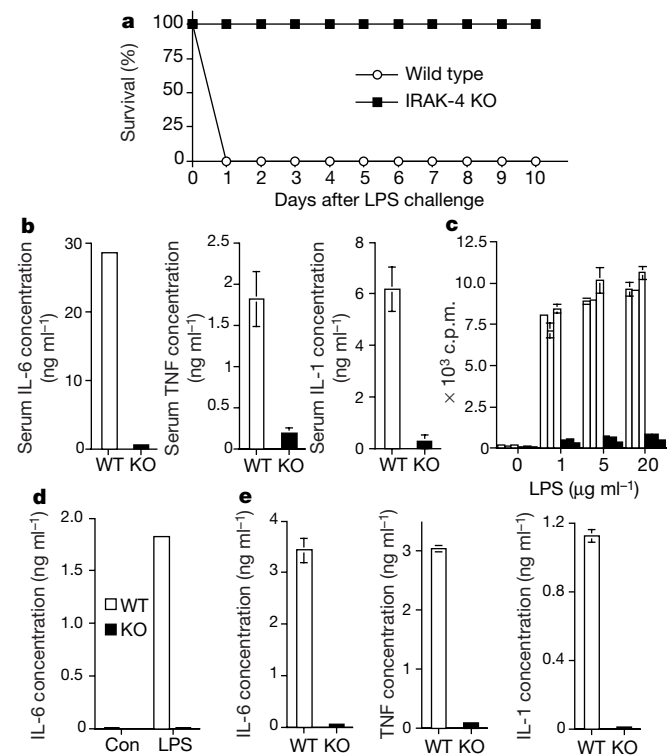


Figure 3 Impaired responses to LPS in *Irak-4*^{-/-} mice and cells. **a**, Resistance of *Irak-4*^{-/-} mice to LPS-induced septic shock. Survival of wild-type (WT) and knockout (KO) mice (6 mice per group) injected with LPS (40 mg kg⁻¹) was monitored daily. **b**, Serum concentrations of IL-6, TNF- α and IL-1 of wild-type and knockout mice (3 per group) at 2 h after LPS injection. **c**, Proliferation of LPS-stimulated purified splenic B cells. Cell proliferation was determined by [³H]thymidine uptake. Data from three experiments are shown. **d**, Induction of IL-6 production in bone marrow-derived macrophages treated for 72 h with 1 μ g ml⁻¹ LPS. Con, untreated control. **e**, Cytokine production by *Irak-4*^{-/-} MEFs treated with 10 μ g ml⁻¹ LPS for 24 h.

and TLR3 signalling also requires IRAK-4. These defects in TLR2, TLR3 and TLR9 signalling were further confirmed in parallel stimulation experiments performed using wild-type and *Irak-4*^{-/-} MEFs (Fig. 4d). Again, a profound impairment of the production of pro-inflammatory cytokines was observed in IRAK-4-deficient cells.

As a lack of IRAK-4 abolished cellular responses to various TLR ligands, we examined the *in vivo* immune responses of *Irak-4*^{-/-} animals to viral and bacterial challenges. IFN- γ and natural killer (NK) cells have important roles in early innate defence against lymphocytic choriomeningitis virus (LCMV). LCMV-induced production of IFN- γ is dependent on IL-12 and IL-18, and IL-18 binds to a receptor that is homologous to IL-1R²⁶. As shown in Fig. 5a, infection of wild-type mice with LCMV normally induces strong production of IFN- γ . However, when *Irak-4*^{-/-} mice were challenged with LCMV, no IFN- γ production could be detected (Fig. 5a). Interestingly, cytolytic activity was intact in IRAK-4-deficient NK cells, as the cytotoxicity of YAC-1 target cells *in vitro* was comparable between wild-type and *Irak-4*^{-/-} NK cells isolated three days after virus infection (Fig. 5b). These results suggest that, although IRAK-4 is required for IFN- γ production by NK cells, the cytotoxicity of NK cells can be supported by compensatory

IFN- α/β or other mechanisms in *Irak-4*^{-/-} animals²⁷.

To investigate the role of IRAK-4 in defence against bacterial infection, we infected wild-type and *Irak-4*^{-/-} mice with 1×10^7 colony-forming units (CFU) of *Staphylococcus aureus* intraperitoneally, and monitored the survival of the mice. As shown in Fig. 5c, whereas 80% of infected wild-type mice survived for at least 10 days, all IRAK-4-deficient animals succumbed to the infection and died within 10 days of inoculation. These results further demonstrate the *in vivo* significance of IRAK-4 in TLR signalling and pathogen-induced immune responses.

The intracellular signalling cascade induced by IL-1R or TLR engagement culminates in the activation of NF- κ B, a transcription factor crucial for inducing the expression of various pro-inflammatory cytokines and mediators. We therefore examined the effect of IRAK-4 deficiency on IL-1-induced NF- κ B activation in *Irak-4*^{-/-} cells. Gel mobility shift assays showed that IRAK-4 is indeed required for IL-1-induced NF- κ B activation but is dispensable for TNF-induced NF- κ B activation (Fig. 6a). I κ B phosphorylation (not shown) and degradation (Fig. 6b) induced by IL-1 was completely inhibited in *Irak-4*^{-/-} macrophages, suggesting that the signalling defect (and thus IRAK-4) lies upstream of the IKK complex. We also assessed whether IRAK-4 was required in macrophages for NF- κ B activation induced by LPS (Fig. 6a, b), poly(I:C) or PGN (Fig. 6b). Of note, delayed and decreased NF- κ B DNA-binding activity and I κ B degradation was observed in LPS-stimulated *Irak-4*^{-/-} macrophages compared with wild-type controls, whereas little or no I κ B degradation occurred in poly(I:C)- or PGN-treated *Irak-4*^{-/-} macrophages (Fig. 6a, b). These results suggest that the absence of IRAK-4 causes significant defects in a variety of TLR signalling pathways.

TIR-containing receptors also activate stress kinases such as c-Jun amino-terminal kinase (JNK). To determine whether the signalling pathway leading to JNK activation was intact in IRAK-4-deficient cells, we treated MEFs with IL-1 or TNF- α , immunoprecipitated JNK, and assayed JNK activity *in vitro* using glutathione S-transferase (GST)-c-Jun as a substrate. Phospho-c-Jun levels were assessed by immunoblotting using an antibody specific for the phosphory-

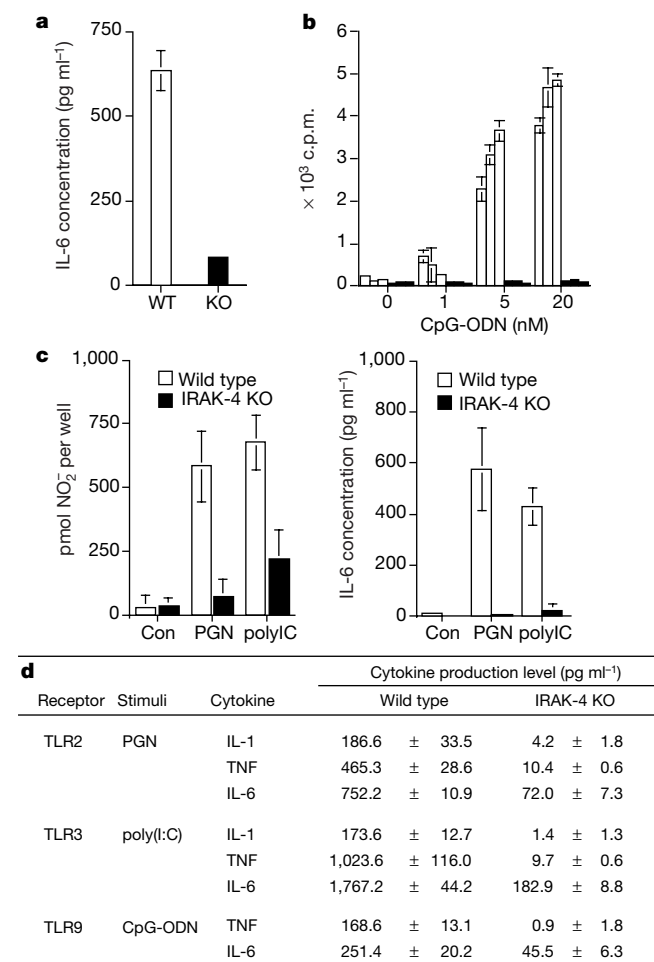


Figure 4 Responses of *Irak-4*^{-/-} mice and cells to TLR2, TLR3 and TLR9 ligands. **a**, Cytokine responses of mice to TLR9 engagement. Serum IL-6 levels were measured 2 h after injection of CpG-ODN plus D-galactosamine. **b**, B-cell proliferation response to CpG-ODN. Data from three experiments are shown. **c**, Inflammatory responses to TLR2 or TLR3 engagement. Con, 50 U ml⁻¹ IFN- γ ; PGN, IFN- γ plus 10 μ g ml⁻¹ peptidoglycan; poly(I:C), 50 U ml⁻¹ IFN- γ plus 100 μ g ml⁻¹ poly(I:C). **d**, Cytokine responses of MEFs to TLR engagement. Wild-type and knockout MEFs were stimulated for 24 h with PGN, poly(I:C), or CpG-ODN (10 μ M). Results shown are the mean $\pm$ s.d. of triplicate determinations for each treatment.

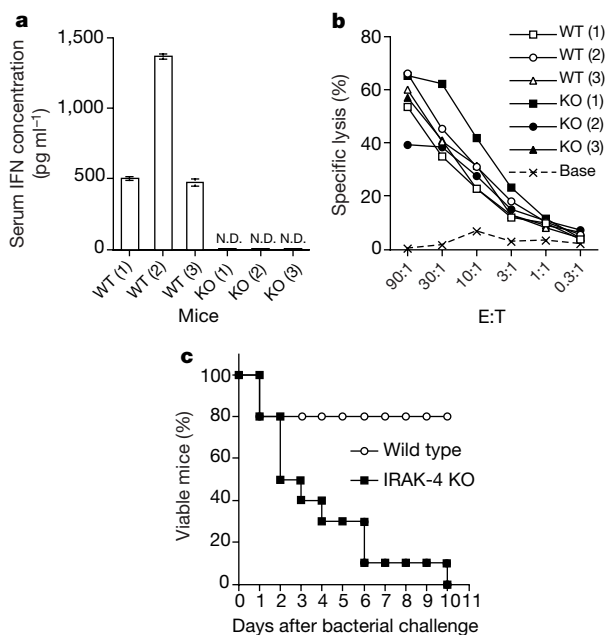


Figure 5 Response of *Irak-4*^{-/-} mice to challenge with LCMV or *S. aureus*. **a**, Serum IFN- γ levels after virus infection. Serum IFN- γ was measured by ELISA on day 3 after infection. N.D., not detectable. **b**, NK cell activity after viral infection. Base, activity of NK cells isolated from an uninfected B6 mouse. **c**, Survival of *Irak-4*^{-/-} mice after bacterial (*S. aureus*) infection. Survival of wild-type and knockout mice (10 per group) was recorded daily. E:T, effector cells versus target cells.

lated form of c-Jun. In lysates of cells lacking IRAK-4, JNK activity induced in response to IL-1 was severely impaired, but that induced by TNF- α treatment was normal (Fig. 6c). We also assessed LPS-induced JNK activity using the same assays. Compared with wild-type controls, *Irak-4*^{-/-} macrophages showed a significant defect in JNK activation in response to LPS stimulation (Fig. 6c). Furthermore, as shown in Fig. 6d, activation of p38 mitogen-activated protein kinase (MAPK) induced by IL-1 (but not by TNF- α) was profoundly inhibited in *Irak-4*^{-/-} cells. These data suggest that IRAK-4 has a critical role in the activation of stress signalling pathways induced by IL-1 and LPS.

We next investigated the position of IRAK-4 in the signalling cascade mediated by TIR-containing receptors. We transfected wild-type and *Irak-4*^{-/-} cells with an NF- κ B-dependent luciferase reporter construct and expression vectors for MyD88, Mal (a MyD88 homologue that has a function in TLR4 signalling)²⁸, or TRAF6. As expected, overexpression of MyD88, Mal or TRAF6

triggered the activation of NF- κ B in wild-type cells, but only overexpression of TRAF6 was able to activate NF- κ B in cells lacking IRAK-4 (Fig. 6e). These results indicate that MyD88 and Mal are upstream mediators requiring IRAK-4 for signal transduction, whereas TRAF6 transmits signals downstream of IRAK-4.

In this study, we have characterized the physiological functions of a new IRAK family member, IRAK-4. Signalling and cellular responses to IL-1 and various TLR ligands are profoundly impaired in *Irak-4*^{-/-} mice, important evidence that IRAK-4 is a principal mediator in these signalling pathways. Mice specifically lacking IRAK-1 exhibit severe but incomplete defects in responses to IL-1, IL-18 and LPS, suggesting that optimal signal transduction involves both IRAK-1 and IRAK-4. Indeed, preliminary evidence has suggested that IRAK-4 associates with IRAK-1 in the signalling complex in a rapid, transient and IL-1-dependent manner¹⁸. As IRAK-4, like IRAK-1, transduces signals between MyD88 and TRAF6, further studies on the functional interactions between

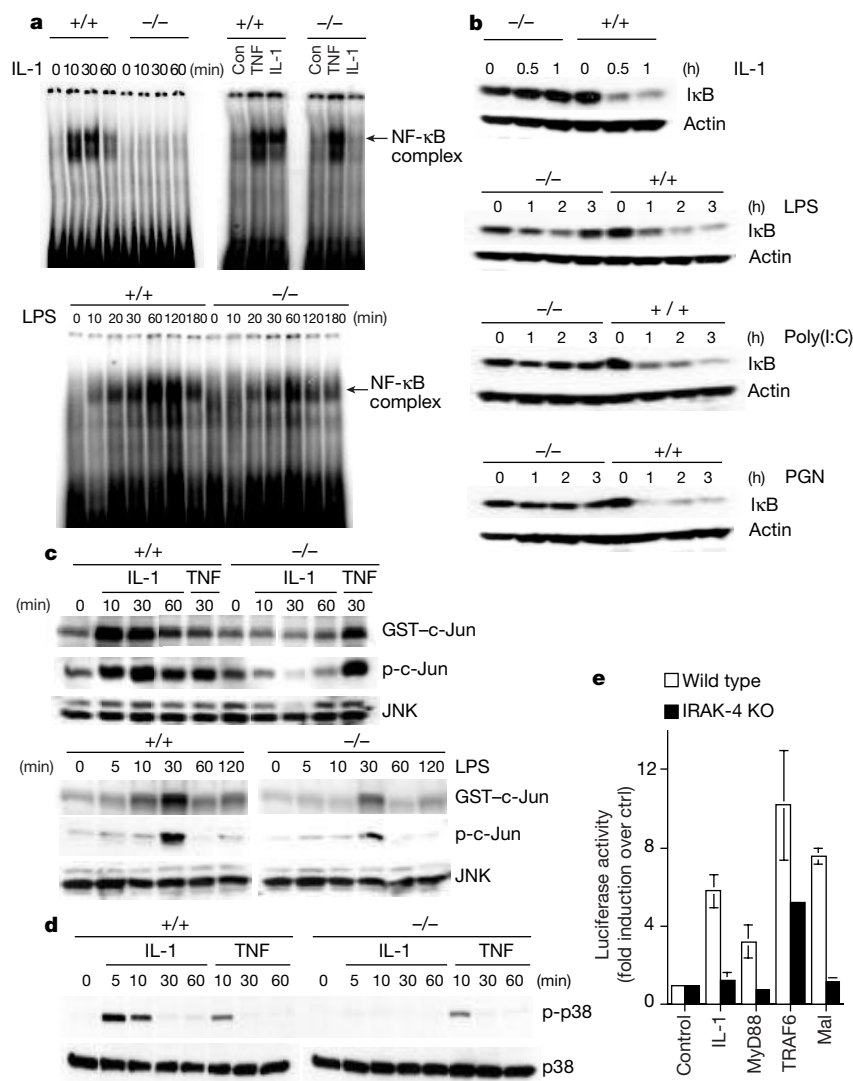


Figure 6 Requirement of IRAK-4 for signalling downstream of IL-1R or TLR engagement. **a**, NF- κ B activation measured by gel mobility shift assay. Wild-type and knockout MEFs were treated with IL-1 for the indicated times (top left), or with 10 ng ml⁻¹ TNF- α or IL-1 for 30 min (top right). Wild-type and knockout bone-marrow-derived macrophages were treated with 1 μ g ml⁻¹ LPS for the indicated times (bottom). For this figure, all time points are in minutes, except for **b**. **b**, Degradation of I κ B protein in bone-marrow-derived macrophages treated with IL-1, LPS (1 μ g ml⁻¹), poly(I:C) (100 μ g ml⁻¹) or PGN (10 μ g ml⁻¹). **c**, JNK activation induced by pro-inflammatory cytokines and LPS. JNK kinase activity was measured as described in Methods. Total cell lysates were

immunoblotted using anti-phospho-c-Jun (JNK activity) and anti-JNK antibodies (loading control). **d**, p38 MAPK activation induced by pro-inflammatory cytokines. Total lysates were immunoblotted to detect phosphorylated p38 MAPK (p-p38) or total p38 MAPK (loading control). **e**, Position of IRAK-4 in the TIR-mediated signalling cascade as determined by transient transfection of wild-type and knockout MEFs with a luciferase reporter and the indicated expression vectors. Control, empty vector, no IL-1 treatment; IL-1, empty vector plus IL-1 treatment; MyD88, TRAF6 and Mal, the relevant expression vectors, no IL-1 treatment.

these IRAK family members will be required to delineate the molecular mechanisms underlying this signalling cascade. □

Methods

Generation of *Irak-4*^{-/-} mice

A genomic DNA clone for IRAK-4 was isolated from a 129/J mouse genomic library using the 5' end of the mouse *Irak-4* complementary DNA as a probe. The targeting construct was designed to disrupt the ATG start codon and to replace part of exon 2 with a PGK-Neo cassette. The targeting plasmid was linearized and transfected into embryonic stem cells of a 129/Ola background (E14 clone). Homologously recombined embryonic stem cell lines were injected into day 3.5 C57BL/6J (B6) blastocysts that were subsequently transferred to CD1 pseudo-pregnant foster mothers, to generate chimaeric progeny. Male chimaeras were backcrossed to C56BL/6J females to produce agouti progeny. Germline transmission in F₁ heterozygous offspring was verified by Southern blot analysis, and F₁ heterozygotes were interbred to obtain homozygous mutant mice.

B-cell purification and thymidine incorporation

Spleens were sieved, washed and lysed using red blood cell (RBC) lysis buffer (Sigma). Splenic leukocytes were washed and resuspended in 3 ml RPMI medium containing 10% FCS. Splenic B lymphocytes were purified by subjecting leukocytes to two rounds of T-cell depletion using Pan T magnetic beads (Dynal), after which cells were washed, enumerated and examined by flow cytometry to assess purity of B cells. The resulting cell suspensions routinely consisted of greater than 80% pure B220⁺ cells. For B-cell proliferation experiments, B cells (1 × 10⁵ per well) were either left untreated, treated with LPS (*E. coli* O111:B4; Sigma), or treated with CpG-ODN (5'-TCGTCGTTTGTCTGTTTGTCTGTT-3'; phosphorothioate stabilized) for 60 h. During the final 18 h, we pulsed cells with 1 µCi [³H]thymidine. Plates were collected and analysed for [³H]thymidine incorporation using a filtermate harvester and a Matrix 96 reader (Packard).

Measurement of IL-6, TNF-α, IL-1 and nitric oxide production

Blood samples were collected from *Irak-4*^{-/-} and control littermates after intraperitoneal injections of IL-1, LPS, or CpG-ODN (5'-TCCATGACGTTCCCTGATGCT-3') at the concentrations and time points indicated in Figs 2–4. Cultures of MEFs or macrophages were stimulated with IL-1, LPS, PGN (*S. aureus*; Fluka), poly(I:C) (Amersham), or CpG-ODN as indicated in the figure legends, and collected after 24 or 72 h, respectively. Measurements of IL-6, TNF-α and IL-1 in serially diluted samples of culture supernatants were performed in triplicate, using murine IL-6, TNF-α and IL-1 enzyme-linked immunosorbent assay (ELISA) kits (R&D Systems) according to the manufacturer's instructions. Nitric oxide production was determined by the Greiss method as described previously²⁹.

Viral and bacterial infections

Mice were infected intravenously with 1 × 10⁶ plaque-forming units of LCMV (Armstrong strain). Sera obtained from LCMV-infected mice on day 3 after infection were quantified for IFN-γ levels by sandwich ELISA using a comparison with recombinant mouse IFN-γ standards. To determine NK cell cytolytic activity, spleen cells of LCMV-infected mice were prepared on day 3 after infection and assayed for lysis of YAC-1 cells in a standard ⁵¹Cr release assay for 4.5 h. Per cent specific lysis was determined from the activity in the culture supernatants according to the formula: (sample c.p.m. – spontaneous c.p.m.)/(total c.p.m. – spontaneous c.p.m.) × 100; where c.p.m. is counts per minute. For bacterial infections, *S. aureus* was cultured in trypticase soy medium, collected, and resuspended in PBS before use. Mice were injected intraperitoneally with 1 × 10⁷ CFU viable *S. aureus*, and the survival of animals was monitored for 10 days.

Gel electrophoresis mobility shift assay

Cells were prepared from MEFs (2 × 10⁶) at various time points after treatment with either 10 ng ml⁻¹ mouse IL-1β (R&D Systems) or 10 ng ml⁻¹ mouse TNF-α (R&D Systems). Nuclear extract (10 µg) was incubated with an end-labelled, double-stranded, NF-κB-specific oligonucleotide probe, and the reaction and gel fractionation were performed according to previously described protocols³⁰.

JNK activity assay

JNK proteins in total cell lysates were immunoprecipitated at 4 °C using anti-JNK polyclonal antibody (C-17, Santa Cruz). Kinase activity in the precipitated fractions was determined using GST-c-Jun as a substrate in the presence of 60 µM [γ-³²P]ATP. The amount of total JNK protein in the immunoprecipitated lysate was determined by western blotting using the polyclonal anti-JNK antibody.

Western blot analysis

Western blots of cell extracts were prepared and incubated with anti-phospho-c-Jun, anti-phospho-p38 MAPK, anti-p38 MAPK, anti-IκB antibodies (all from New England Biolabs), anti-β-actin antibody (Fisher Scientific), or anti-IRAK-4 polyclonal antibody (raised against peptide CLHEKKNRRPDIAKVQQLQEMSA, which corresponds to the C-terminus of mouse IRAK-4). Protein bands were visualized using secondary horseradish peroxidase (HRP)-conjugated antibodies and the ECL detection system (Amersham Pharmacia).

NF-κB reporter assays

Transient transfections were performed using Elam promoter (NF-κB-dependent) luciferase reporter plus pCMV-MyD88, pCMV-Mal, or pCMV-TRAF6, or an empty

vector as the control. pCMV-lacZ was included in each transfection to normalize results for transfection efficiency, and luciferase and β-galactosidase activities were measured as described previously³⁰.

Received 12 December 2001; accepted 13 March 2002.

Published online 31 March 2002, DOI 10.1038/nature736.

- O'Neill, L. The Toll/interleukin-1 receptor domain: a molecular switch for inflammation and host defence. *Biochem. Soc. Trans.* **28**, 557–563 (2000).
- Akira, S., Takeda, K. & Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nature Immunol.* **2**, 675–680 (2001).
- Kimbrell, D. A. & Beutler, B. The evolution and genetics of innate immunity. *Nature Rev. Genet.* **2**, 256–267 (2001).
- Wasserman, S. A. Toll signaling: the enigma variations. *Curr. Opin. Genet. Dev.* **10**, 497–502 (2000).
- Wesche, H., Henzel, W. J., Shillinglaw, W., Li, S. & Cao, Z. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity* **7**, 837–847 (1997).
- Medzhitov, R. *et al.* MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol. Cell* **2**, 253–258 (1998).
- Adachi, O. *et al.* Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143–150 (1998).
- Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
- Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. & Goeddel, D. V. TRAF6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
- Lomaga, M. A. *et al.* TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev.* **13**, 1015–1024 (1999).
- Cao, Z., Henzel, W. J. & Gao, X. IRAK: a kinase associated with the interleukin-1 receptor. *Science* **271**, 1128–1131 (1996).
- Muzio, M., Ni, J., Feng, P. & Dixit, V. M. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science* **278**, 1612–1615 (1997).
- Wesche, H. *et al.* IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. *J. Biol. Chem.* **274**, 19403–19410 (1999).
- Kanakaraj, P. *et al.* Interleukin (IL)-1 receptor-associated kinase (IRAK) requirement for optimal induction of multiple IL-1 signalling pathways and IL-6 production. *J. Exp. Med.* **187**, 2073–2079 (1998).
- Thomas, J. A. *et al.* Impaired cytokine signaling in mice lacking the IL-1 receptor-associated kinase. *J. Immunol.* **163**, 978–984 (1999).
- Shelton, C. A. & Wasserman, S. A. *pelle* encodes a protein kinase required to establish dorsoventral polarity in the *Drosophila* embryo. *Cell* **72**, 515–525 (1993).
- Shiu, S. H. & Blecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
- Li, S., Strelow, A., Fontana, E. J., Wesche, H. IRAK-4—a novel member of the IRAK family with the properties of an IRAK-kinase. *Proc. Natl Acad. Sci. USA* (in the press).
- Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
- Underhill, D. M. *et al.* The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).
- Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).
- Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
- Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
- Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
- Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).
- Parnet, P., Garka, K. E., Bonnett, T. P., Dower, S. K. & Sims, J. E. IL-1Rrp is a novel receptor-like molecule similar to the type I interleukin-1 receptor and its homologues T1/ST2 and IL-1RAcP. *J. Biol. Chem.* **271**, 3967–3970 (1996).
- Orange, J. S. & Biron, C. A. Characterization of early IL-12, IFN-α/β, and TNF effects on antiviral state and NK cell responses during murine cytomegalovirus infection. *J. Immunol.* **156**, 4746–4756 (1996).
- Fitzgerald, K. A. *et al.* Mal (MyD88-adaptor-like) is required for Toll-like receptor-4 signal transduction. *Nature* **413**, 78–83 (2001).
- Nguyen, L. T. *et al.* TRAF2 deficiency results in hyperactivity of certain TNFR1 signals and impairment of CD40-mediated responses. *Immunity* **11**, 379–389 (1999).
- Bonnard, M. *et al.* Deficiency of T2K leads to apoptotic liver degeneration and impaired NF-κB-dependent gene transcription. *EMBO J.* **19**, 4976–4985 (2000).

Acknowledgements

We thank C. Richardson, S. Pownall, A. Suzuki, Y. Horie, M. Bonnard and S. Der for reagents and discussions. We also thank E. Lye, T. Horacek and S. Sugimoto for technical assistance, and M. Saunders for scientific editing. This work was supported in part by the National Cancer Institute of Canada (NCIC), with funds from the Terry Fox Run (W.-C.Y.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.-C.Y. (e-mail: wyeh@amgen.com).

Cells compete for Decapentaplegic survival factor to prevent apoptosis in *Drosophila* wing development

Eduardo Moreno^{*†}, Konrad Basler^{*} & Ginés Morata[†]

^{*} Institut für Molekularbiologie, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[†] Centro de Biología Molecular CSIC-UAM, Cantoblanco, 28049 Madrid, Spain

During the growth of *Drosophila* imaginal discs a process called 'cell competition'¹ eliminates slow-proliferating but otherwise viable cells. We report here that cell competition requires the function of the *brinker* (*brk*) gene, whose expression is normally repressed by Decapentaplegic (Dpp) signalling^{2–4} but is upregulated in slow-growing *Minute/+* cells. Excess *brk* expression activates the c-Jun amino-terminal kinase pathway, which in turn triggers apoptosis in these cells. We propose that slow-proliferating cells upregulate *Brk* levels owing to a disadvantage in competing for, or in transducing, the Dpp survival signal. This sequence of events might represent a general mechanism by which weaker cells are eliminated from a growing population, and might serve as a method of controlling cell number and optimizing tissue fitness and hence organ function.

The regulation of organ size is a largely unexplored field of developmental biology^{5,6}. An intriguing aspect of size control in *Drosophila* was discovered in experiments in which the division rate of wing cells was altered by using *Minute* (*M*) mutations¹. These are dominant mutations, defective in the production of ribosomal proteins⁷. Heterozygous *M/+* animals are of normal size but their development is delayed with respect to wild-type flies. Their developmental delay is due to a lower proliferation rate of *M/+* cells than that of wild-type (hereafter *M*⁺) cells. Clones of *M*⁺ cells are larger than *M/+* clones initiated at the same time in the same disc and can comprise up to 90% of the anterior or posterior compartment¹. This indicates that there is a size control mechanism that can cope with differential division rates: to compensate for the

faster growth of the *M*⁺ clone, the *M/+* cells contribute less to the compartment than they would have in the absence of the *M*⁺ clone. The process by which *M/+* cells are eliminated by the presence of faster-dividing *M*⁺ cells is referred to as 'cell competition'^{1,8,9}. Cell competition is also implicated in the elimination of slow-dividing cells mutant for genes involved in cellular growth^{10–13}.

We have followed the time course of the elimination of slow-proliferating *M/+* cells in *M*⁺ wing discs by inducing *M/+* clones and fixing the discs at various times after induction. After heat-shock-induced FRT/FLP recombination, two types of marked clones can be scored (Fig. 1); one (*M/+*) containing only a single

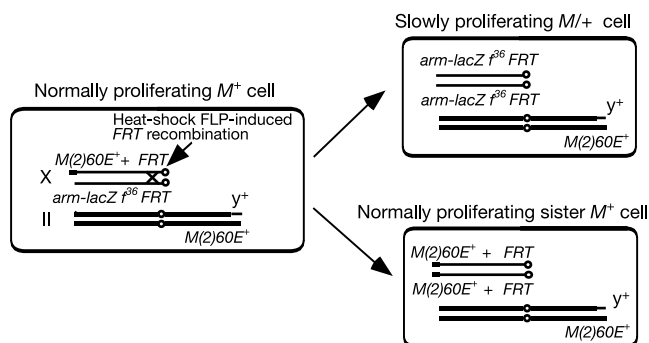


Figure 1 Mitotic recombination method for inducing marked clones with one or three doses of the *M(2)60E* gene (see Methods for further details). Recombination in the X chromosome (arrow) gives rise after mitosis to two sister cells: one of them (top) has lost one copy of *M(2)60E* but still retains the other copy on the second (II) chromosome. This cell divides more slowly than surrounding cells and is homozygous for the marker mutant *f^{36a}*, which alters bristle differentiation and makes its progeny scorable in the adult cuticle. It also carries two doses of the *arm-lacZ* insert, which allows us to distinguish the clone generated by this cell because of the stronger *lacZ* activity. The sister cell (bottom) will originate the 'twin' clone. These cells contain three doses of *M(2)60E*, carry the wild-type allele of *forked* and have lost the *arm-lacZ* insertion; they therefore do not contain β -Gal activity.

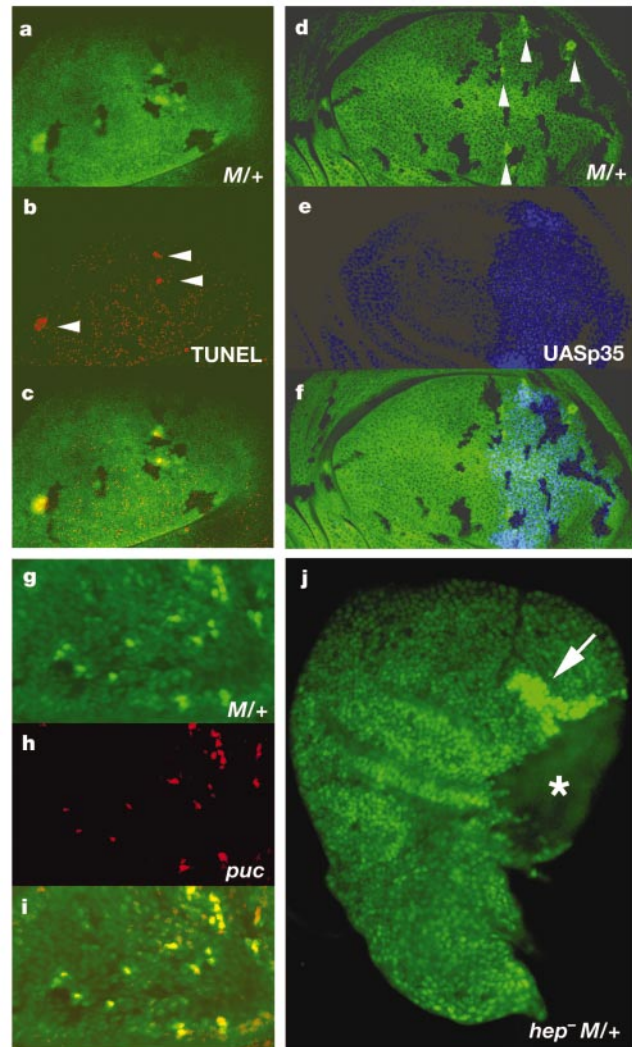


Figure 2 Cell competition is mediated by JNK-dependent apoptosis. **a–c**, Apoptosis during cell competition. **a**, Wing disc (green stain) with *M*⁺ (lack of green staining) and slow-growing *M/+* clones (bright green). **b**, TUNEL staining. **c**, Merge. Several slow-growing *M/+* clones show TUNEL staining (arrowheads in **b**). **d–f**, p35 rescues the elimination of slow-growing cells. **d**, *M/+* (bright green, arrowheads) and *M*⁺ clones (lack of green staining) induced in a disc expressing the apoptosis inhibitor p35 in the P compartment. **e**, P compartment labelled blue with anti-Engrailed antibody. **f**, Merge. Note the survival of *M/+* clones in the P compartment but not in the A compartment. **g–i**, *puc* is upregulated in slow-growing cells undergoing cell competition. **g**, *M*⁺ (absence of green staining) and *M/+* (bright green) clones. **h**, *puc-lacZ* expression (red). **i**, Merge showing *puc* upregulation in *M/+* cells. **j**, Cell-competition-mediated apoptosis is dependent on JNK signalling. *M/+*; *hep⁷⁵* double-mutant clone (arrow, bright green) and *M*⁺ clone (dark, marked with asterisk). Note that although the *M/+*; *hep⁷⁵* mutant clone is not eliminated by apoptosis, it is smaller than the wild-type *M*⁺ clone.

copy, and the twin (M^+) containing three copies of the $M(2)60E$ gene. By 24 h after clone initiation both $M/+$ and M^+ clones are readily detected (Fig. 2a, g), but after 24 h the $M/+$ clones are clearly smaller than their twins, and their frequency decreases. At 48 h after clone initiation, only the twin M^+ clones are found (Fig. 3h). These results agree with previous observations in the adult cuticle showing that $M/+$ clones are eliminated after 24 h (ref. 8).

Because $M/+$ cells are viable ($M/+$ animals survive to adulthood), their elimination might be caused by apoptosis, triggered by the presence of faster-dividing M^+ cells in the same primordium. We stained discs containing $M/+$ clones for TdT-mediated dUTP nick end labelling (TUNEL)¹⁴ and found evidence for apoptosis associated with the clones (Fig. 2a–c). We also found that cell competition can be reduced by the expression of baculovirus p35 protein, which inhibits apoptosis¹⁵. $M/+$ clones were induced in $hh-Gal4/UAS-p35$ discs in which p35 is expressed only in P-compartment cells. At the time when $M/+$ clones have been eliminated from the A compartment, many $M/+$ clones are present in the P compartment (Fig. 2d–f).

The c-Jun N-terminal kinase (JNK) pathway has been shown to mediate apoptosis in wing cells¹⁶, so we tested whether it mediates elimination of $M/+$ cells. JNK signalling can be monitored by the expression levels of *puckered* (*puc*), a gene coding for dual-specificity phosphatase that forms a negative feedback loop by down-regulating the activity of JNK¹⁷. We find that *puc* is activated in $M/+$ cells (Fig. 2g–i). Moreover, the elimination of $M/+$ cells can be

prevented if the JNK pathway is blocked; $M/+$ cells survive if they are mutant for *hemipterous* (*hep*) (Fig. 2j), which codes for the mitogen-activated protein kinase kinase of the JNK pathway¹⁸. We conclude that $M/+$ cells are eliminated by the autonomous up-regulation of JNK signalling, which programmes these cells to undergo apoptosis.

A key factor promoting growth in the wing disc is Decapentaplegic (Dpp), a member of the superfamily of transforming growth factor- β growth factors^{6,19}. The Dpp signalling pathway seems to activate genes necessary for cell proliferation¹⁹. An important control element in the Dpp pathway is *brinker* (*brk*), a transcriptional repressor²⁰, which prevents the response of Dpp target genes and is itself negatively regulated by the Dpp pathway^{2–4}. Cell clones lacking the Dpp receptor Thick veins (Tkv) upregulate *brk* expression, divide slowly and are eliminated^{4,19}, except in the region that normally contains high *brk* levels¹⁹. The elimination of *tkv* mutant cells is presumably due to *brk* upregulation, because *tkv brk* double-mutant clones are not eliminated⁴. Interestingly, we observe (Fig. 3a–d, and data not shown) that $M/+$ clones survive better in the domain in which *brk* is expressed at high levels, the same region as that in which *tkv* clones are not eliminated^{4,19}. We have also examined the survival of $M/+$ cells in a complementary experiment, inducing large overgrowing M^+ clones in $M/+$ wings (data not shown). These clones are induced early in development so that they fill most of the corresponding compartment. The remaining $M/+$ cells appear preferentially in the lateral region of the discs and in the

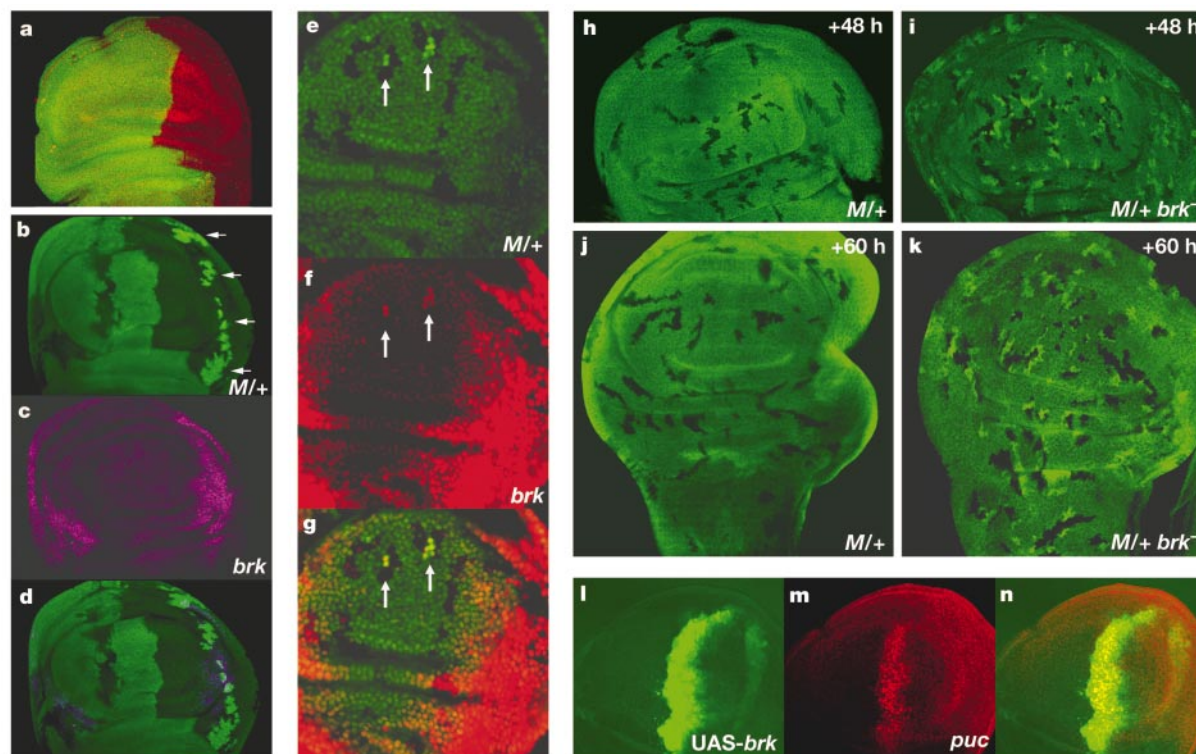


Figure 3 Cell competition is associated with, and caused by, *brk* upregulation. **a–d**, Induction of large numbers of $M/+$ and M^+ clones in the P compartment with the use of UAS-FLP and *hh-Gal4* (see Methods). Essentially all cells in the P compartment undergo recombination and produce equal numbers of slow-growing $M/+$ (bright green) and wild-type M^+ clones (dark) (see Methods). **a**, Disc in which all the P-compartment cells (labelled red with anti-Engrailed antibody) are M^+ . Slow-growing $M/+$ cells have been eliminated by cell competition. **b–d**, Slow-growing $M/+$ cells survive longer in lateral regions where *brk* is expressed. **b**, One disc in which some islands of $M/+$ cells remain (arrows). **c**, The same disc stained for *brk* expression (pink). **d**, Merge. Note the good correlation between the *brk* domain and the zone where $M/+$ cells survive longer. **e–g**, Upregulation of *brk* in slow-growing $M/+$ clones. **e**, $M/+$ clones (arrows, bright green) located away from the *brk* domain (**f**, red). Note in **f** the gain of *brk* activity (arrows) by the $M/+$ cells. **g**, Merge of panels **e** and **f**. **h–k**, Removal of *brk* activity allows the survival of $M/+$ clones. **h, j**, Wild-type M^+ (black) and slow-growing $M/+$ (bright green) clones induced 48 h (**h**) and 60 h (**j**) before fixation. The slow-growing $M/+$ clones have been eliminated. **i, k**, $M/+$ clones that lack *brk* function (bright green) survive 48 h (**i**) and 60 h (**k**) after induction (compare with **h, j**). **l–n**, *brk* overexpression induces *puc*. **l**, *dpp-Gal4/UAS-brk* discs with UAS-GFP staining (green). **m**, *puc* upregulation revealed by *puc-lacZ* staining. **n**, Merge. *puc* (red) is activated in the region with *brk* overexpression (green).

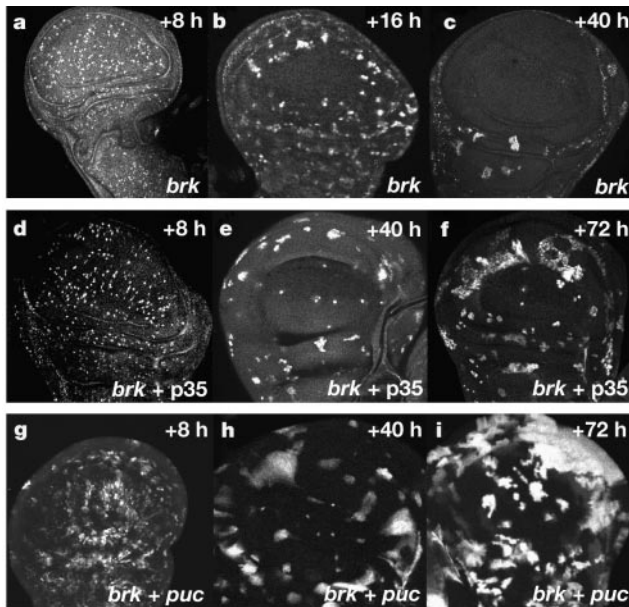


Figure 4 Brk induces JNK-dependent apoptosis. **a–c**, Clones of *brk*-expressing cells stained 8, 16 and 40 h after clone induction. Note the disappearance of the clones from the centre. Clones localized in the lateral region, where *brk* is normally expressed at high levels, survive (see clones in **c**). **d–f**, Clones of cells coexpressing *brk* and p35, initiated 8, 40 and 72 h before fixation. The presence of p35 allows the survival of *brk*-expressing cells in the centre of the disc, although their proliferation is impaired. **g–i**, Clones of cells coexpressing *brk* and *puc*, stained 8, 40 and 72 h after clone induction. The hyperactivation of *puc* blocks JNK signalling and allows the survival of *brk*-expressing cells in the centre of the disc.

medial and proximal margin of the adult wings, corresponding to the zones of high requirement for *brk* activity^{2–4}.

The observations above indicate that, as with *tkv* clones, *M/+* cells might acquire inappropriately high or ectopic *brk* expression and that this upregulation, in regions where cells normally have low or no *brk* activity, triggers their elimination. We assayed *brk* activity in *M/+* clones induced in *M*⁺ wing discs and indeed observed ectopic *brk* expression (Fig. 3e–g). We tested this possibility by

Table 1 Recovery of *M/+* clones in the adult notum structures

Genotype	Number of animals examined	Number of <i>f</i> ³⁶ clones per notum	Number of <i>f</i> ³⁶ clones per abdomen
<i>T(1;2)sc^{S2/y} f^{36a}</i>	23	3/23 = 0.13	147/23 = 6.4
<i>T(1;2)sc^{S2/y} brk^{Δh} f^{36a}</i>	27	220/27 = 8.14	183/25 = 7.3

genetic methods and analysed the survival of *M/+* clones that are mutant for a *brk*-null allele. By 48 h after clone initiation all *M/+* clones have disappeared, whereas *brk; M/+* clones induced in parallel are still present (Fig. 3h, i). Even after 60 h of clone induction most of the *brk; M/+* clones are still detectable (Fig. 3j, k). The enhanced survival of *brk; M/+* clones can also be demonstrated in the differentiated adult cuticle, where both *M/+* and *brk; M/+* cells were labelled with the marker *forked* (*f*^{36a}) (Table 1): *brk; M/+* clones are recovered with a 60-fold greater frequency than *M/+* clones. As a control we measured the frequency of *f*^{36a} clones in the abdomen, where there is no cell competition¹, and found the same frequency for the two genotypes. In the wing disc the average size of the *brk; M/+* clones is smaller than that of *M*⁺ clones. In a control experiment with the same *f*³⁶ marker, the average size of normal clones in the notum initiated 48–72 h after egg laying is 3.6 bristles per clone (*n* = 197), whereas it is 1.3 for a sample of *brk; M/+* clones induced in parallel (*n* = 107). Thus, the loss of *brk* activity prevents elimination of *M/+* clones but does not restore normal proliferation rate.

The observation that the loss of *brk* function strongly reduces cell competition indicated that the upregulation of *brk* seen in *M/+* cells (Fig. 3f) might activate JNK signalling and consequently apoptosis. Indeed, we find that mere overexpression of *brk* in the centre of the disc induces *puc* expression (Fig. 3l–n). This led us to test whether inappropriately high *brk* levels alone induce JNK-dependent apoptosis. We first observed that *brk*-expressing clones disappear from the centre of the disc (Fig. 4a–c) and survive only in lateral regions, where cells normally express high levels of *brk*. The elimination of *brk*-expressing clones is caused by apoptosis because coexpression of the apoptosis suppressor protein p35 results in the survival of these clones (Fig. 4d–f). Moreover, the elimination of *brk*-expressing cells requires JNK signalling, because they survive if *puc*

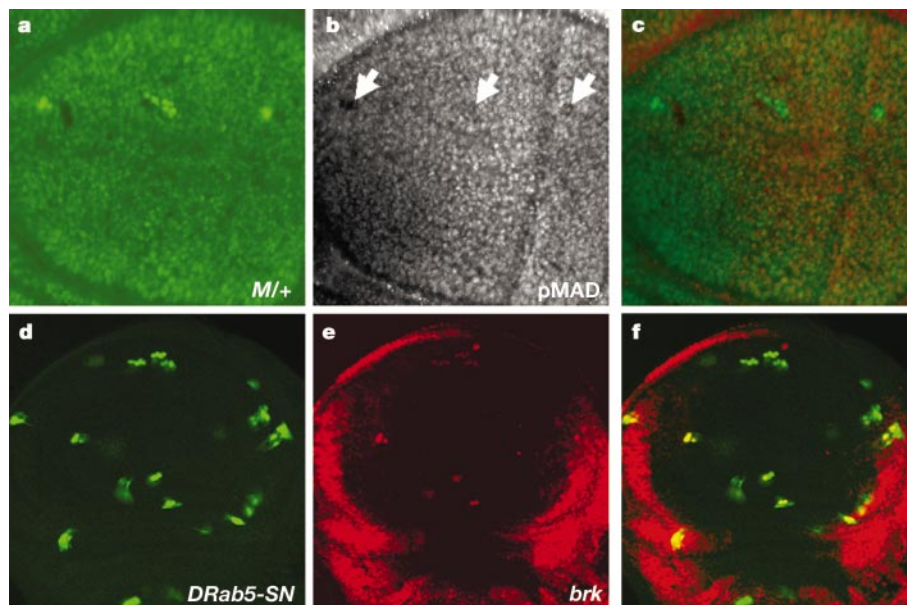


Figure 5 Decrease in Dpp signalling during cell competition and during endocytic blockage. **a–c**, *M/+* clones show lower pMAD levels than neighbour cells. **a**, *M/+* clones (bright green) after 24 h of clone induction. **b**, pMAD levels are lower in *M/+* cells (arrows). **c**, Merge of panels **a** and **b**. **d–f**, Clones of cells expressing a dominant-negative version of *Drab5* (green) show ectopic *brk* activity.

expression is induced concomitantly (Fig. 4g–i). High levels of *puc* are known to repress JNK activity¹⁷. These results indicate that elevated levels of *brk* cause programmed cell death, but only in regions where Dpp is required for growth and cell survival, and where *brk* expression is normally low or absent. It is unknown how ectopic *brk* expression elicits JNK activity. The JNK pathway might be triggered indirectly as a consequence of a positional mis-specification of *brk*-expressing cells. Alternatively, Brk might directly repress a negative regulator of the JNK pathway.

Our results indicate a simple model that might explain the principal features of cell competition. The basis of cell competition is that it recognizes the presence of cells with different proliferation rates. Subsequently, it triggers a mechanism to eliminate the slower-growing ones. Here we have identified three critical elements of this mechanism: upregulation of *brk* expression, followed by activation of the JNK pathway, and then programmed cell death. Because it seems that *brk* expression is the first of these sequential events, we propose that cell competition is initiated by a debilitation of the Dpp pathway in wing cells that have a lower division rate than their neighbours. Indeed, we observe that Dpp transduction levels, as indicated by Mad phosphorylation²¹, are lower in *M/+* cells than in their neighbouring *M⁺* cells (Fig. 5a–c). Internalization by endocytosis might be necessary to carry Dpp to an endosomal compartment where the signal transduction takes place²². The GTPase Rab5, which is required for the formation of endocytic vesicles and fusion with early endosomes, has been implicated in Dpp²³ signalling. We observe that expression of a dominant-negative form of DRab5 causes an upregulation of *brk* levels (Fig. 5d–f), suggesting that endocytosis of Dpp is required before signal transduction and that competition for the Dpp signal might take place at the level of ligand internalization. Hence *M/+* cells might capture less Dpp ligand—perhaps simply because they are metabolically less active than surrounding cells—causing reduced Dpp transduction and the upregulation of *brk*. Higher *brk* levels in turn reduce proliferation further and trigger JNK-dependent apoptosis, resulting in the elimination of such cells. The link between Dpp signalling and cell competition might explain why the latter does not occur in the dorsal abdomen¹, a body region in which Dpp does not have a function in regulating growth and patterning²⁴.

Mechanisms based on competition for limiting amounts of survival factors have been postulated previously to explain the regulation of cell number during development^{5,25}. Cells unable to obtain sufficient quantities of survival factors undergo apoptosis and are eliminated. Our model is formally similar: we propose that one of the factors that imaginal disc cells compete for is the Dpp signal. Our model is not only applicable to cells that divide more slowly than their neighbours: any cell that, for whatever reason, transduces the Dpp signal less efficiently would undergo a similar process of *brk* upregulation followed by JNK activation and apoptosis. Cell competition might be a mechanism that serves not only to ensure the correct size and shape of the *Drosophila* wing but also to recognize and eliminate less adapted or weaker cells of developmental primordia. If conserved throughout evolution, cell competition might be a general mechanism by which animals maximize tissue fitness to optimize and maintain organ function. □

Methods

Generation of marked clones

For *M/+* clones in *M⁺* discs and adults, females of genotype *T(1;2)sc⁵²FRT18A/FM6/CyO* were crossed with either *arm-lacZ FRT18A:hs-FLP* or *hs-GFP hs-FLP FRT18A* males to induce marked clones in discs, or to *y w f^{36a} FRT18A/FM6 hs-FLP* males to induce clones scorable in adult derivatives. The key element of the method is that the *T(1;2)sc⁵²* carries a wild-type copy of *M(2)60E*, the gene that codes for the ribosomal protein L19 (RpL19) (ref. 26), to the first chromosome and therefore the second chromosome of the translocation is defective for *M(2)60E⁺*. A recombination event (see Fig. 1) between the first chromosomes of females *M(2)60E⁺/ywf^{36a}*; *DfM(2)60E/+* produces a cell that loses a copy of *M(2)60E⁺* and therefore becomes heterozygous for the retarding *Df(2)M60E/+* condition.

Larvae of genotype *T(1;2)sc⁵²FRT18A/arm-lacZ FRT18A; hs-FLP; hh-Gal4/UAS-p35* were subjected to a heat shock to produce *M/+* and sibling *M⁺* clones expressing high levels of p35 protein if situated in the posterior compartment but no p35 if situated in the anterior compartment.

Monitoring JNK signalling in *M/+* clones: larvae of the genotype *hs-GFP hsFLP FRT18A/T(1;2)sc⁵²FRT18A; puc⁶⁰⁹/+* were subjected to a 20-min heat shock at 37 °C and harvested 36 h later.

To prevent JNK signalling in *M/+* clones, larvae of genotype *hep^{r75} hs-GFP hsFLP FRT18A/T(1;2)sc⁵²FRT18A* were subjected to a heat shock and fixed 72 h later.

For *M/+* clones in *M⁺* discs in a *hh-Gal4/UAS-FLP* background, larvae of genotype *T(1;2)sc⁵²FRT18A/arm-lacZ FRT18A; UAS-FLP; hh-Gal4/+* were generated. In wing discs of this genotype, the high levels of FLP recombinase driven by the *hh-Gal4* driver ensured that all the cells in the posterior compartment underwent mitotic recombination, forming two genetically different populations, *M/+* and *M⁺*. Marked *M/+* and twin *M⁺* clones were produced in equal numbers.

To monitor *brk* expression in *M/+* clones, larvae of the genotype *hs-GFP hsFLP FRT18A/T(1;2)sc⁵²FRT18A; brk-lacZ/+* were subjected to a heat shock and fixed 36 h later. *brk-lacZ* is a transgene comprising the regulatory region of the *brk* gene and is situated on the third chromosome (B. Müller and K.B., unpublished observations).

For the induction of *brk*; *M/+* clones, larvae of genotype *T(1;2)sc⁵²FRT18A/yw brk^{sh} arm-lacZ FRT18A; hs-FLP/+* and adults of genotype *T(1;2)sc⁵²FRT18A/yw brk^{sh} f^{36a} FRT18A; hs-FLP/+* were generated. *brk^{sh} M/+* clones were marked by strong *lacZ* expression in discs and by the forked phenotype in adults.

Gain of *brinker* expression in adults and discs: to induce marked clones of *brk*-expressing cells, *yw hs-FLP f^{36a}; abx > f⁺ > Gal4-lacZ/+; UAS-y/UAS-brk* larvae were generated and subjected to a heat shock. *brk*-expressing cells were marked by gain of *lacZ* activity in the discs and by the *y⁺* and *f^{36a}* markers in adult derivatives. To induce marked *brk*-expressing clones that concomitantly expressed p35 or *puc*, *UAS-p35* or *UAS-puc* transgenes were introduced into the above genotypes.

Gain of dominant-negative *DRab5* expression in discs: to induce marked clones in discs, *yw hs-FLP; brk-lacZ/+; UAS-GFP; actin > CD2 > Gal4/UAS-5SN* larvae were generated and subjected to a heat shock.

Staining with antibodies

Imaginal discs were immunostained by using the standard procedures for confocal microscopy. The specific antibodies used were rabbit anti-LacZ (Cappel), monoclonal anti-Myc (Babco), rat anti-Brk (a gift from G. Campbell) and rabbit anti-pMad (a gift from P. ten Dijke). TUNEL assays were performed as described in ref. 26.

Received 30 November 2001; accepted 11 February 2002.

- Morata, G. & Ripoll, P. Minutes: mutants of *Drosophila* autonomously affecting cell division rate. *Dev. Biol.* **42**, 211–221 (1975).
- Minami, M., Kinoshita, N., Kamoshida, Y., Tanimoto, H. & Tabata, T. *brinker* is a target of Dpp in *Drosophila* that negatively regulates Dpp-dependent genes. *Nature* **398**, 242–246 (1999).
- Jazwinska, A., Kirov, N., Wieschaus, E., Roth, S. & Rushlow, C. The *Drosophila* gene *brinker* reveals a novel mechanism of Dpp target gene regulation. *Cell* **96**, 563–573 (1999).
- Campbell, G. & Tomlinson, A. Transducing the Dpp morphogen gradient in the wing of *Drosophila*: regulation of Dpp targets by *brinker*. *Cell* **96**, 553–562 (1999).
- Conlon, I. & Raff, M. Size control in animal development. *Cell* **96**, 235–244 (1999).
- Lawrence, P. A. Morphogens: how big is the big picture? *Nature Cell Biol.* **3**, 151–154 (2001).
- Lambertson, A. The Minute genes in *Drosophila* and their molecular functions. *Adv. Genet.* **38**, 69–134 (1998).
- Simpson, P. Parameters of cell competition in the compartments of the wing of *Drosophila*. *Dev. Biol.* **69**, 182–193 (1979).
- Simpson, P. & Morata, G. Differential mitotic rates and pattern of growth in compartments in the *Drosophila* wing. *Dev. Biol.* **85**, 299–308 (1981).
- Böhni, R. et al. Autonomous control of cell and organ size by CHICO, a *Drosophila* homolog of vertebrate IRS1–4. *Cell* **97**, 865–875 (1999).
- Johnston, L. A., Prober, D. A., Edgar, B. A., Eisenman, R. N. & Gallant, P. *Drosophila myc* regulates cellular growth during development. *Cell* **98**, 779–790 (1999).
- Prober, D. & Edgar, B. *ras1* promotes cellular growth in the *Drosophila* wing. *Cell* **100**, 435–446 (2000).
- Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. & Edgar, B. A. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93**, 1183–1193 (1998).
- Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119**, 493–501 (1992).
- Hay, B. A., Wolff, T. & Rubin, G. M. Expression of baculovirus p35 prevents cell death in *Drosophila*. *Development* **120**, 2121–2129 (1994).
- Adachi-Yamada, T., Fujimura-Kamada, K., Nishida, Y. & Matsumoto, K. Distortion of proximodistal information causes JNK-dependent apoptosis in *Drosophila* wing. *Nature* **400**, 166–169 (1999).
- Martin-Blanco, E. et al. *puckered* encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.* **12**, 557–570 (1998).
- Glise, B., Bourbon, H. & Noselli, S. *hemipterous* encodes a novel *Drosophila* MAP kinase kinase, required for epithelial cell sheet movement. *Cell* **83**, 451–461 (1995).
- Burke, R. & Basler, K. Dpp receptors are autonomously required for cell proliferation in the entire developing *Drosophila* wing. *Development* **122**, 2261–2269 (1996).
- Zhang, H., Levine, M. & Ashe, H. L. *Brinker* is a sequence-specific transcriptional repressor in the *Drosophila* embryo. *Genes Dev.* **15**, 261–266 (2001).
- Tanimoto, H., Itoh, S., ten Dijke, P. & Tabata, T. Hedgehog creates a gradient of DPP activity in *Drosophila* wing imaginal discs. *Mol. Cell* **5**, 59–71 (2000).
- Tsukazaki, T., Chiang, T. A., Davidson, A. F., Attisano, L. & Wrana, J. L. SARA, a FYVE domain protein that recruits Smad2 to the TGF- β receptor. *Cell* **95**, 779–791 (1998).
- Entchev, E. V., Schwabedissen, A. & Gonzalez-Gaitan, M. Gradient formation of the TGF- β homolog Dpp. *Cell* **103**, 981–991 (2000).
- Struhl, G., Barbash, D. A. & Lawrence, P. A. Hedgehog acts by distinct gradient and signal relay

mechanisms to organise cell type and cell polarity in the *Drosophila* abdomen. *Development* **124**, 2155–2165 (1999).

25. Raff, M. C. Social controls on cell survival and cell death. *Nature* **356**, 397–400 (1992).

26. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).

27. Milán, M., Campuzano, S. & García-Bellido, A. Developmental parameters of cell death in the wing disc of *Drosophila*. *Proc. Natl Acad. Sci. USA* **94**, 5691–5696 (1997).

Acknowledgements

We thank T. Adachi-Yamada, G. Campbell, M. Calleja, A. Cantarero, S. González-Crespo, M. Gonzalez-Gaitan, E. Hafen, A. Jazwinska, E. Martin-Blanco, B. Müller, E. Sánchez-Herrero, M. Suzanne and P. ten Dijke for fly stocks and antibodies; R. González for technical assistance; S. Campuzano for help with the TUNEL technique; C. Sánchez for help with confocal microscopy; and M. Calleja, O. Gerlitz, E. Hafen, G. Hausmann, P. Lawrence, F. Martin, B. Müller, D. Nellen, E. Sánchez-Herrero, H. Stocker and members of our laboratories for discussions. E.M. was supported by a postdoctoral fellowship from the Comunidad Autónoma de Madrid and the Swiss National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.B. (e-mail: basler@molbio.unizh.ch) or G.M. (e-mail: gmorata@cbm.uam.es).

SWAP-70 is a guanine-nucleotide-exchange factor that mediates signalling of membrane ruffling

Masahiro Shinohara*, Yoh Terada*†, Akihiro Iwamatsu†, Azusa Shinohara†, Naoki Mochizuki†, Maiko Higuchi§, Yukiko Gotoh§, Sayoko Ihara*, Satoshi Nagata*†, Hiroshi Itoh*, Yasuhisa Fukui* & Rolf Jessberger||

* Department of Applied Biological Chemistry, Faculty of Agricultural and Life Science, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan
† Central Laboratories for Key Technology, Kirin Brewery, 1-13-5 Fukuura, Kanazawa-ku, Yokohama City, Kanagawa 236-0004, Japan
‡ Department of Pathology, International Medical Center of Japan, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan
§ Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
|| Mount Sinai School of Medicine, The Carl C. Icahn Institute for Gene Therapy and Molecular Medicine, 1425 Madison Avenue, Box 1496, New York, New York 10029-6574, USA

Phosphoinositide-3-OH kinase (PI(3)K), activated through growth factor stimulation, generates a lipid second messenger, phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃)^{1–5}. PtdIns(3,4,5)P₃ is instrumental in signalling pathways that trigger cell activation, cytoskeletal rearrangement, survival and other reactions. However, some targets of PtdIns(3,4,5)P₃ are yet to be discovered^{1–7}. We demonstrate that SWAP-70, a unique signalling protein^{8–10}, specifically binds PtdIns(3,4,5)P₃. On stimulation by growth factors, cytoplasmic SWAP-70, which is dependent on PI(3)K but independent of Ras, moved to cell membrane rearrangements known as ruffles. However, mutant SWAP-70 lacking the ability to bind PtdIns(3,4,5)P₃ blocked membrane ruffling induced by epidermal growth factor or platelet-derived growth factor. SWAP-70 shows low homology

with Rac-guanine nucleotide exchange factors (GEFs), and catalyses PtdIns(3,4,5)P₃-dependent guanine nucleotide exchange to Rac. SWAP-70-deficient fibroblasts showed impaired membrane ruffling after stimulation with epidermal growth factor, and failed to activate Rac fully. We conclude that SWAP-70 is a new type of Rac-GEF which, independently of Ras, transduces signals from tyrosine kinase receptors to Rac.

SWAP-70 was originally identified as a protein involved in B-cell activation and heavy-chain immunoglobulin class switching^{8,9}. It is abundantly expressed in activated B lymphocytes and in immature mast cells. SWAP-70 localizes to the cytoplasm in these cells, but in B cells it also translocates to the nucleus after appropriate stimulation, which is consistent with a role in immunoglobulin class switching. This translocation, the association of SWAP-70 with the B-cell antigen receptor complex and motifs such as its pleckstrin homology (PH) domain, indicated that SWAP-70 might be involved in signal transduction. The protein has not been detected in the nucleus of a large variety of other cell types or tissues, but the presence of low amounts of SWAP-70 in the cytoplasm of other cell types has not been excluded^{8–10}. Mice that are deficient in SWAP-70 are phenotypically healthy but their B cells are hypersensitive to γ -irradiation, show compromised CD40 signalling, and show impaired switching to the immunoglobulin- ϵ (IgE) class¹¹. The signalling pathways within which SWAP-70 acts and its mechanism of action are poorly understood.

PI(3)K is activated immediately upon stimulation of a variety of cells with growth or differentiation factors. Its major product, PtdIns(3,4,5)P₃, is a lipid second messenger that is important in various signalling pathways, including triggering of vesicle trans-

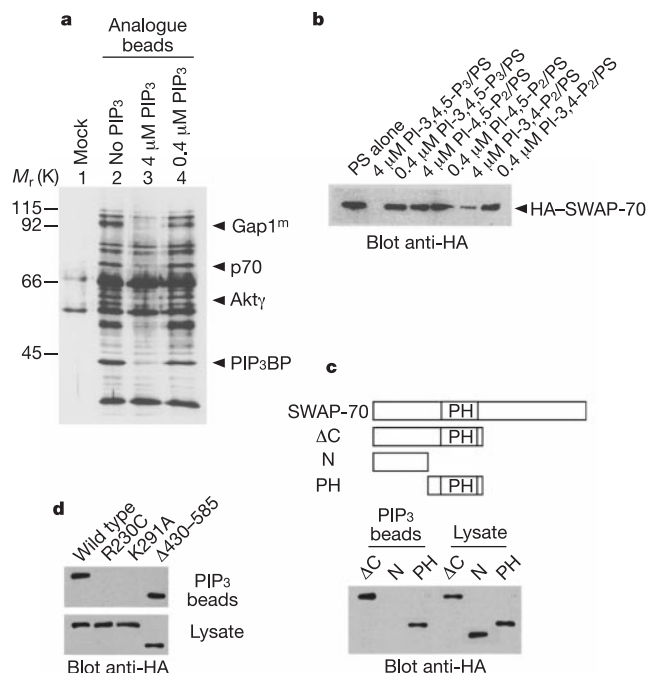


Figure 1 SWAP-70 binds PtdIns(3,4,5)P₃. **a**, Bovine brain lysate was incubated with (0.4 or 4 μ M) or without PtdIns(3,4,5)P₃ (PIP₃), and mixed with PtdIns(3,4,5)P₃ analogue beads¹². Bead-bound proteins were analysed by SDS–polyacrylamide gel electrophoresis and silver staining. Mock, beads without PtdIns(3,4,5)P₃; PIP₃BP, PtdIns(3,4,5)P₃-binding protein. **b**, Specificity of binding. Lysates prepared from 293T cells expressing HA-SWAP-70 were incubated with phosphoinositides as indicated (PS, phosphatidylserine) and then with PtdIns(3,4,5)P₃ beads. Anti-HA antibody detected bead-bound SWAP-70. **c**, PH domain is required. HA-SWAP-70 deletion mutants were expressed in 293T cells and tested for binding to PtdIns(3,4,5)P₃. Δ C, mutant with deleted C-terminal region; N, N-terminal region of SWAP-70. **d**, As in **c** but with HA-SWAP-70 bearing different point mutations in the PH domain. Δ 430–585, amino acids 430–585 deleted.

† Present addresses: Molecular Medicine Laboratories, Institute for Drug Discovery Research, Yamanouchi Pharmaceutical Co. Ltd, Tsukuba, Ibaraki 305-8585, Japan (Y.T.); Department of Structure Analysis, National Cardiovascular Center Research Institute, 5-7-4 Fujishirodai, Suita, Osaka 565-8565, Japan (N.M.); Laboratory of Molecular Biology, National Cancer Institute, National Institute of Health, Bethesda, Maryland 20892-4255, USA (S.N.).

port, cytoskeletal rearrangement and cell survival^{1–5}. PI(3)K can be activated by binding to various proteins such as tyrosine-phosphorylated growth-factor receptors, tyrosine-phosphorylated non-receptor cytoplasmic proteins, $\beta\gamma$ subunits of trimeric G proteins, or activated Ras⁶. Several laboratories have explored the targets of PI(3)K and have identified some PtdIns(3,4,5)P₃-binding proteins that might be substrates for PtdIns(3,4,5)P₃ modification^{6,7}. Considering the functional multiplicity of PI(3)K, a significant number of unknown PtdIns(3,4,5)P₃ targets probably remain to be discovered⁷. We therefore searched for PtdIns(3,4,5)P₃-binding proteins by affinity chromatography on PtdIns(3,4,5)P₃ analogue beads¹².

Total cell lysate from bovine brain was mixed with the beads, and the bound proteins were analysed. Free PtdIns(3,4,5)P₃ enhanced the specificity of binding, which can be competed for by further increased PtdIns(3,4,5)P₃. This approach allowed us to identify several known PtdIns(3,4,5)P₃-binding proteins such as Akt/protein kinase B, PtdIns(3,4,5)P₃-binding protein and Gap1^m (refs 12,13,14,15). However, a protein with a relative molecular mass of 70,000 (*M_r* 70K) seemed to be unknown (Fig. 1a). PtdIns(3,4)P₂ and PtdIns(4,5)P₂ did not block binding of the *M_r* 70K protein to the PtdIns(3,4,5)P₃ beads, indicating that binding is specific for PtdIns(3,4,5)P₃. Peptide sequences of the purified protein showed almost perfect identity to mouse and human SWAP-70, suggesting that p70 is its bovine homologue (Supplementary Information Fig. S1). Although SWAP-70 expression in brain has not been reported,

this tissue also contains a large number of mast cells¹⁶ that might have facilitated the detection of SWAP-70 in the total brain lysate.

SWAP-70 was expressed in 293T cells and subjected to the PtdIns(3,4,5)P₃-binding assay. SWAP-70, tagged with the influenza virus haemagglutinin (HA) epitope, bound efficiently to the PtdIns(3,4,5)P₃ beads (Fig. 1b). Preincubation with other polyphosphoinositides such as PtdIns(3,4)P₂ and PtdIns(4,5)P₂ did not effectively block the binding. Mutational analysis determined that SWAP-70 binds to PtdIns(3,4,5)P₃ as long as it contains the PH domain. Two point mutations in the PH domain, the R230C mutant analogous to that found in Btk of mice with X-linked immunodeficiency¹⁷, and K291A, both abolished the binding of PtdIns(3,4,5)P₃ (Fig. 1c, d). Thus, the PH domain of SWAP-70 is essential for PtdIns(3,4,5)P₃ binding.

To determine the subcellular localization of SWAP-70, the protein was fused to green fluorescent protein (GFP) and was expressed in NIH 3T3 cells. GFP–SWAP-70 was localized mainly in the cytosol (Fig. 2a; Supplementary Information Fig. S2). We examined the behaviour of SWAP-70 upon stimulation with growth factor, which also activates PI(3)K (ref. 5). GFP–SWAP-70 was localized together with actin filaments at membrane ruffles 3 min after stimulation with platelet-derived growth factor (PDGF); 30 min after stimulation, membrane ruffling declined and SWAP-70 returned to the cytosol, suggesting that the translocation of SWAP-70 is reversible (Supplementary Information Fig. S3). The PI(3)K inhibitor

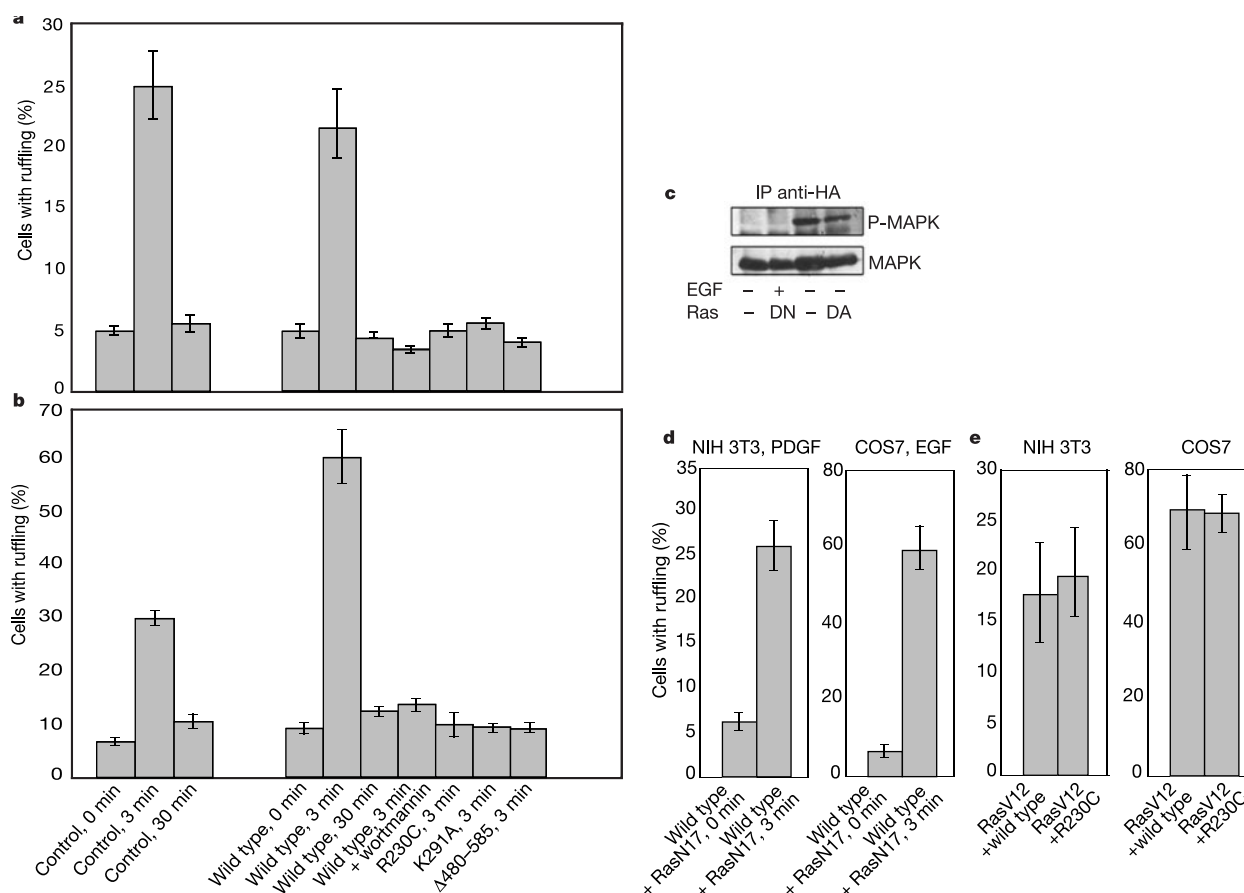


Figure 2 SWAP-70 translocates to membrane ruffles. **a**, **b**, SWAP-70 translocates to membrane ruffles after growth-factor treatment. NIH 3T3 (**a**) or COS7 (**b**) cells, expressing wild-type or mutant GFP–SWAP-70, were stimulated with PDGF (10 ng ml⁻¹) (**a**) or EGF (100 ng ml⁻¹) (**b**) and stained for F-actin. In one experiment, cells were pretreated for 30 min with 100 nM wortmannin. **c**, Dominant-negative Ras blocks the Ras pathway completely. COS7 cells were transfected with or without expression vectors for RasV12 and RasN17. Cells were treated with EGF (100 ng ml⁻¹) for 3 min and the phosphorylation

of MAP kinase was analysed by western blotting with anti-phospho-MAP kinase antibody. The lower panel shows expression of MAP kinase detected by western blotting with anti-MAP kinase antibody. **d**, RasN17 does not block membrane ruffling in NIH 3T3 or COS7 cells. IP, immunoprecipitation. **e**, Dominant-negative SWAP-70 does not inhibit membrane ruffling induced by RasV12. Cells with membrane ruffles were scored and normalized to all fluorescent cells. Means for three independent experiments are shown; bars indicate s.d.

wortmannin completely inhibited this reaction. Mutant SWAP-70 protein lacking the ability to bind $\text{PtdIns}(3,4,5)\text{P}_3$ not only failed to localize with actin filaments but also blocked membrane ruffling, as did a mutant lacking the carboxy-terminal region. Similar effects were observed in COS7, Chinese hamster ovary and Madin–Darby canine kidney cells stimulated with epidermal growth factor (EGF), insulin and hepatocyte growth factor, respectively (Fig. 2b; Supplementary Information Fig. S2b; and data not shown). These results indicate that SWAP-70 might be required for the formation of membrane ruffles.

Ras-activated PI(3)K is known to induce membrane ruffling through Vav, Sos and Tiam1, that is, $\text{PtdIns}(3,4,5)\text{P}_3$ -dependent Rac-GEFs^{18–20}. However, PI(3)K can also be activated to induce membrane ruffling independently of Ras by binding to tyrosine-phosphorylated growth-factor receptors or to proteins that have been tyrosine-phosphorylated by growth-factor receptors^{21–23}. To test whether SWAP-70 signalling is dependent on Ras, dominant-negative Ras was expressed together with SWAP-70 and membrane ruffling was tested. Expression of Ras was readily detectable (data not shown). However, the dominant-negative Ras did not affect membrane ruffling (Fig. 2c; Supplementary Information Fig. S2c). Under identical conditions, dominant-negative Ras completely inhibited the activation of MAPK after treatment with EGF in COS7 cells, indicating that the activity of endogenous Ras was completely blocked (Fig. 2d). These results confirm that a growth-factor signalling pathway triggering actin rearrangement independently of the activation of Ras is present in NIH 3T3 or COS7 cells. Expression of constitutively activated Ras resulted in an induction of membrane ruffling. However, flattening of the cells was observed and membrane ruffling was seen on the edge of the cells, which was somewhat different from ruffling seen after stimulation with growth factors, suggesting that constitutively activated Ras does not completely mimic the effect of stimulation with growth factor. Mutant SWAP-70 that lacks $\text{PtdIns}(3,4,5)\text{P}_3$ -binding activity did not inhibit Ras-induced membrane ruffling (Fig. 2e; Supplementary Information Fig. S2d). This contrasts with membrane ruffling generated

by stimulation with PDGF or EGF (Fig. 2a, b), and also shows that SWAP-70 transduces signals from growth-factor receptors independently of Ras.

Examination of the SWAP-70 amino-acid sequence revealed that the C-terminal portion of SWAP-70 has limited similarity to Dbl homology (DH) domains. The putative DH domain of SWAP-70 is 18.1%, 18.7% and 23.2% identical to that of Vav1, Sos1 and Tiam1, respectively^{18–20}. By comparison, the DH domain of Vav1 is 27.2% and 24.6% similar to that of Tiam1 and Sos1, respectively (Supplementary Information Fig. S4). However, the DH domain of SWAP-70 is flanked at its amino terminus by the PH domain. This differs from other Rac-GEFs, which have C-terminal PH domains, and underscores the uniqueness of SWAP-70. The GEF activity of SWAP-70 was analysed *in vitro* by monitoring the release of [^3H]GDP from purified glutathione *S*-transferase (GST)–Rac1. The rate of [^3H]GDP release was slightly enhanced by the presence of SWAP-70 (Fig. 3a). Although it has been reported that $\text{PtdIns}(3,4,5)\text{P}_3$ alone stimulates the dissociation of GDP from Rac²⁴, $\text{PtdIns}(3,4,5)\text{P}_3$ was not effective in releasing GDP. However, the addition of both SWAP-70 and $\text{PtdIns}(3,4,5)\text{P}_3$ greatly enhanced GEF activity (Fig. 3a). In contrast, SWAP-70 did not show GEF activity on RhoA (Fig. 3b) or Cdc42 (Fig. 3c). An interaction between SWAP-70 and Rac1 was further illustrated by the specific binding of SWAP-70 to Rac1 in absence of Mg^{2+} , but not to RhoA or Cdc42. SWAP-70 bound weakly to GDP-loaded Rac1 but not to GTP- γS -loaded Rac1 (Fig. 3d). This supports the idea that SWAP-70 is a Rac-specific GEF. The R230C mutant lacking $\text{PtdIns}(3,4,5)\text{P}_3$ -binding activity exhibited wild-type-like GEF activity, but it was not enhanced by the addition of $\text{PtdIns}(3,4,5)\text{P}_3$ (Supplementary Information Fig. S5a). Phosphatidylserine, PtdIns , $\text{PtdIns}(3)\text{P}$, $\text{PtdIns}(4)\text{P}$, $\text{PtdIns}(4,5)\text{P}_2$ and $\text{PtdIns}(3,4)\text{P}_2$, which do not bind to SWAP-70 as seen in dot-blot analysis, also did not enhance the GEF activity (Supplementary Information Fig. S5b). Comparing the GEF activities of SWAP-70 and Sos revealed very similar levels of activity (Supplementary Information Fig. S5c). The GEF activity of an N-terminal deletion mutant that had retained

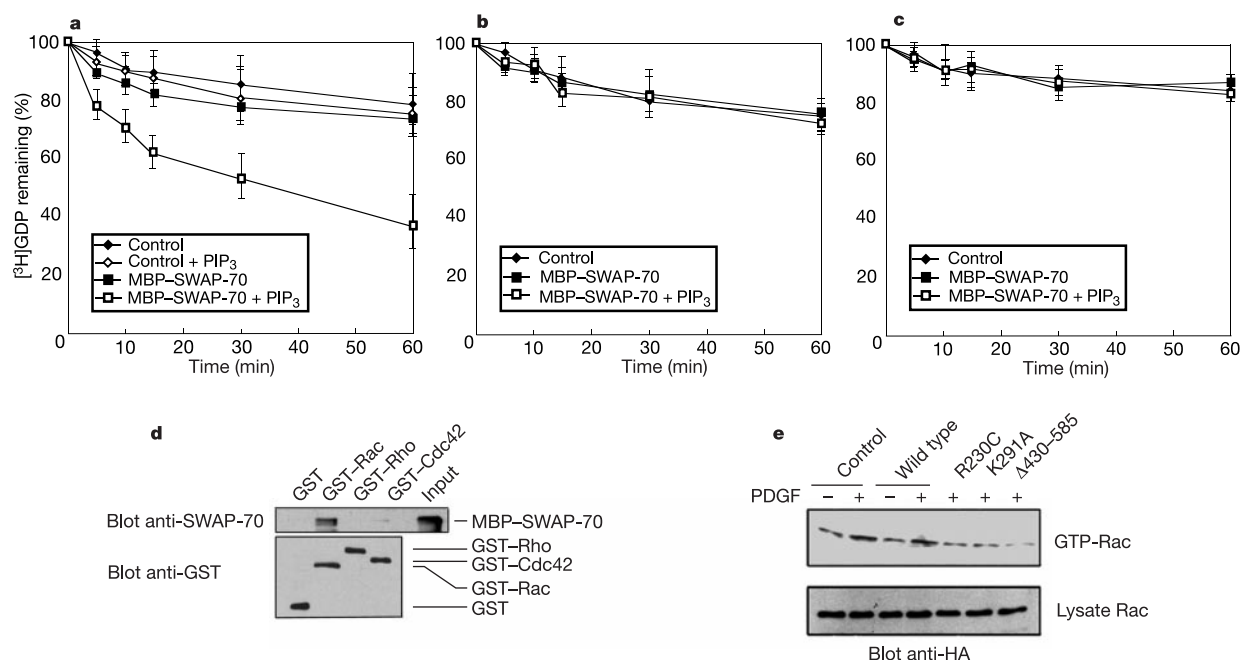


Figure 3 SWAP-70 is a $\text{PtdIns}(3,4,5)\text{P}_3$ -dependent Rac specific GEF. **a–c**, GDP-release assays. [^3H]GDP-loaded Rac1 (**a**), RhoA (**b**) or Cdc42 (**c**) was incubated with MBP or MBP-SWAP-70 with or without $\text{PtdIns}(3,4,5)\text{P}_3$ (PIP_3) (1 μM). **d**, Binding of SWAP-70 to Rac1. Bead-bound GST–Rho GTPases were incubated with MBP–SWAP-70 (no Mg^{2+}). Complexes were analysed by immunoblotting with anti-SWAP-70 (top) or anti-GST-

antibody (bottom). **e**, Inhibition of Rac1 by SWAP-70 mutants. NIH 3T3 cells expressing HA–Rac1 and a wild-type or a mutant GFP–SWAP-70, as indicated, were treated with (+) or without (–) PDGF. Lysates were incubated with GST–CRIB and detected with anti-HA antibody (top); bottom, cell lysates used.

only the DH domain was even higher than that of wild-type SWAP-70 (data not shown). This indicates that the DH domain alone is sufficient for GEF activity and that it might be regulated by interaction of the DH domain with other domains. A deletion mutant that lacks the C-terminal region inhibited membrane ruffling, indicating that it acted dominant-negatively (Fig. 2a). These results indicate that SWAP-70 is a PtdIns(3,4,5)P₃-dependent GEF for Rac and that this activity is correlated with the formation of membrane ruffles. Overexpression of SWAP-70 in COS7 cells resulted in an increase in activated Rac1 concentration, suggesting that SWAP-70 does indeed function as Rac-GEF *in vivo* (data not shown). Direct evidence for the activation of Rac1 *in vivo* was obtained by a pull-down assay specific for activated Rac (Fig. 3e). Only the wild type, but not any of the three SWAP-70 mutants, activated Rac1. Moreover, the dominant-negative C-terminal deletion mutant inhibited Rac1 activation. In contrast, no effect of SWAP-70 on Cdc42 activation was observed (data not shown). Because many of the conserved residues typical of known GEFs are substituted in SWAP-70, and the position of the PH domain is different in other Rac-GEFs, we suggest that SWAP-70 is a new type of Rac-GEF.

To determine further the role of SWAP-70 *in vivo* in membrane ruffling induced by growth factors, kidney fibroblasts (or embryonic fibroblasts that yielded similar results) were cultured from wild-type or SWAP-70-deficient mice¹¹. For ease of direct comparison with the earlier experiments (Figs 2 and 3), we chose fibroblasts although they express only low levels of cytoplasmic SWAP-70. SWAP-70 was detectable in wild-type cells but was absent from SWAP-70-deficient cells (data not shown). SWAP-70-deficient fibroblasts tended to show more stress fibres, and typical lamellipodia were rarely seen. The cells were less motile, forming tighter colonies than wild-type cells (data not shown). Wild-type cells

responded to stimulation by EGF with membrane ruffling in more than 50% of the cells after 5–10 min (Fig. 4a, b). Wortmannin blocked this reaction. In contrast, membrane ruffling was significantly impaired in SWAP-70-deficient cells, as only up to 20% showed membrane ruffling (Fig. 4a, b). In these cells, stimulation with EGF was comparable to that in wild-type cells as judged by tyrosine phosphorylation of the EGF receptor and of other signalling molecules such as MAP kinase or Shc (Supplementary Information Fig. S6). Microinjection of an expression vector for SWAP-70 restored the ruffling deficiency to nearly the wild-type level (Fig. 4c; Supplementary Information Fig. S7). We also noted that high expression levels of SWAP-70 seem to inhibit ruffling in primary wild-type fibroblasts and (mildly) in NIH 3T3 cells. Taken together, these results confirm that SWAP-70 is important in the efficient induction of membrane ruffling in primary fibroblasts. As seen after the expression of mutant SWAP-70 (Fig. 3e), numbers of EGF-stimulated SWAP-70-deficient fibroblasts were significantly lower in Rac1 activation (Fig. 4d). SWAP-70 deficiency had no effect if ruffling was induced through PI(3)K-independent signals such as thrombin (Supplementary Information Fig. S8). Similarly, expression of the dominant-negative mutant R230C in NIH 3T3 cells did not block PI(3)K-independent ruffling induced by 12-O-tetradecanoylphorbol-13-acetate (data not shown). Taken together, these results indicate that SWAP-70 is a PI(3)K-dependent Rac-GEF, which receives signals from growth-factor receptors independently of Ras.

SWAP-70-deficient mice are vital and fertile, and no obvious phenotype other than deficient B-cell activation and dysfunctional mast cell biology have been reported^{11,31}. Indeed, SWAP-70-deficient kidney cells still underwent membrane ruffling to some extent. These observations suggest that there might be alternative, SWAP-70-independent, pathways to activate Rac, perhaps through

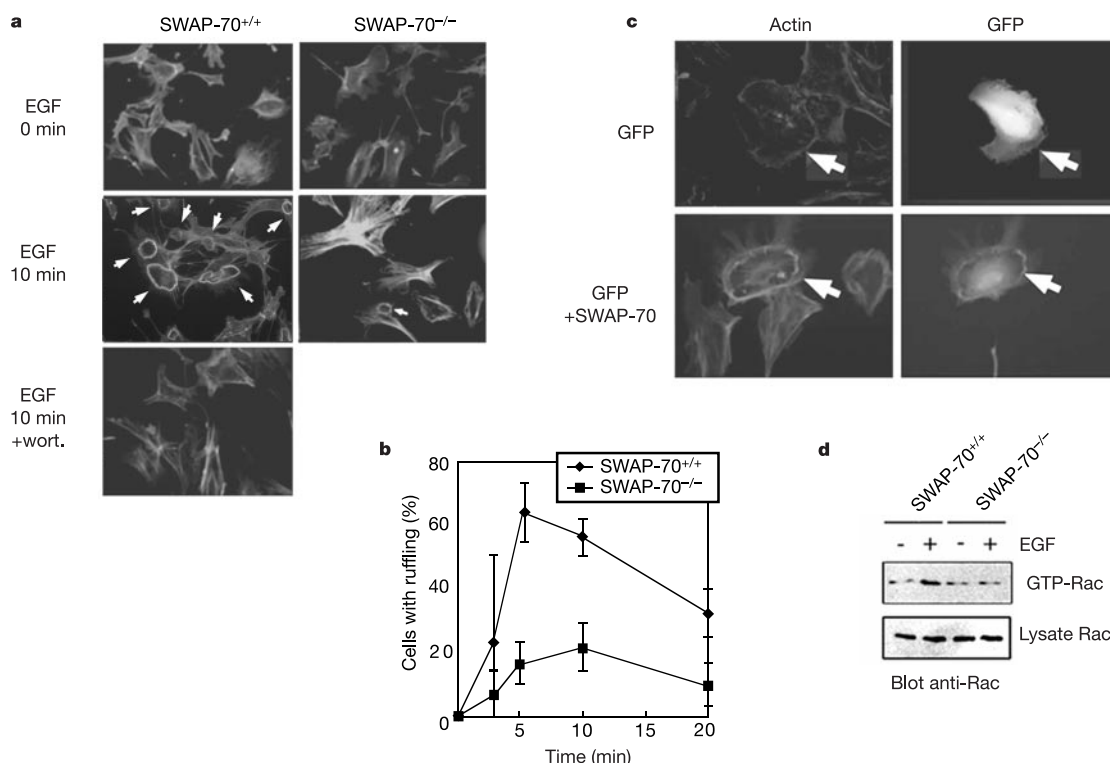


Figure 4 Impaired membrane ruffling in SWAP-70-deficient primary fibroblasts.

a, Membrane ruffling induced by EGF (50 ng ml⁻¹). Cells were stained for F-actin. Arrows indicate cells with membrane ruffling. Wort., wortmannin. **b**, Scores of cells with membrane ruffling. **c**, Microinjection of an expression vector for SWAP-70 restores membrane ruffling. Scores of membrane ruffles in microinjected cells are shown.

d, SWAP-70 regulates Rac activation *in vivo*. SWAP-70 wild-type (SWAP-70^{+/+}) or SWAP-70-deficient (SWAP-70^{-/-}) cells were treated with (+) or without (-) EGF. GTP-Rac1, precipitated from lysates with bead-bound GST-CRIB (top), or Rac1 in the lysate (bottom), was detected by immunoblotting with anti-Rac1 antibody.

EGF- or PDGF-triggered activation of Ras. This is reminiscent of mice deficient in the Eps8 activator component of the Rac-GEF complex Eps8/E3b1/Sos-1 (ref. 25). PI(3)K- and Ras-dependent ruffling is impaired in the mutant, yet the mice are phenotypically healthy. Thus, indications exist *in vivo* for redundancies in Rac-dependent signalling of cytoskeletal rearrangements. □

Methods

Plasmids

KIAA0640 complementary DNA was provided by T. Nagase (Kazusa DNA Research Institute). For GFP-SWAP-70 expression in mammalian cells, cDNA of SWAP-70 was subcloned into pEGFP C1 (Clontech). Point mutations were induced by the QuickChange site-directed mutagenesis method (Stratagene). For the SWAP-70 (R230C and K291A) mutants, 5'-AAAGCTGGAGTGAATGTTGGTTTGTAAACCCAA-3' and 5'-GATAAGACTTTTGAATCAGTGCAGATGCGAAGAAGAACAGGAGTGGATTCAA-3' were used as sense primers, respectively. For deletion mutants, cDNAs coding for different segments were inserted into pEGFP C1. pEFBOS HA-RasN17, pEFBOS HA-RasV12 and pEFBOS HA-Rac were gifts from K. Kaibuchi. pCMV HA-MAPK was provided by Y. Gotoh. cDNA of the GEF domain of human Sos1 was cloned by reverse-transcriptase-mediated polymerase chain reaction.

Protein purification and peptide sequence analysis

SWAP-70 purification from bovine brain with the use of PtdIns(3,4,5)P₃-APB beads and amino-acid sequencing were done as described^{12,26}.

Transfection

Transfection of COS7 cells was performed by electroporation. Cells were suspended in 500 µl of K-PBS (30.8 mM NaCl, 120.7 mM KCl, 8.1 mM Na₂HPO₄, 1.46 mM KCl, 5 mM MgCl₂) containing 20 µg DNA and were pulsed with a Cell-porator (Gibco BRL) at 800 µF capacitance and 240 V. Transfection of NIH 3T3 cells was done by electroporation as described above or by lipofection with LipofectAMINE (Gibco-BRL). Transfection of 293T cells was done as described previously²⁷. Introduction of the DNA to cultivated kidney cells was done by microinjection as described previously²⁸ or by lipofection.

Fluorescence microscopy

Cells were stained for F-actin with tetramethylrhodamine-isothiocyanate-conjugated phalloidin as described previously²⁸. Cells were examined by confocal fluorescence microscopy (Olympus IX-70).

GDP release assay for Rac, binding of SWAP-70 to Rac, and Rac activation assay

The GDP release assay was performed as described²⁹. Binding of SWAP-70 to Rho GTPases was performed as described previously³⁰. For the Rac1 (or Cdc42) activation assay, Rac1 (or Cdc42) was expressed in NIH 3T3, COS7 or primary kidney cells together with wild-type or mutant SWAP-70. Cells were treated for 5 min with or without PDGF or EGF, washed with ice-cold PBS containing 5 mM MgCl₂ lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 1% (v/v) Triton X-100, 150 mM NaCl, 10% (v/v) glycerol, 2 mM dithiothreitol, 10 mM MgCl₂) and centrifuged for 10 min at 15,000g at 4°C. Supernatants were incubated for 30 min at 4°C with GST-Cdc42/Rac interactive binding (CRIB) (from PAK1), prebound to glutathione-Sepharose beads. Beads were washed, and bound material was analysed by SDS-polyacrylamide gel electrophoresis and immunoblotting.

Received 9 October 2001; accepted 11 February 2002.

1. Toker, A. & Cantley, L. C. Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* **387**, 673–676 (1997).
2. Downward, J. Role of phosphoinositide-3-OH kinase in Ras signaling. *Adv. Second Mess. Phosphoprot. Res.* **31**, 1–10 (1997).
3. Martin, T. F. Phosphoinositide lipids as signaling molecules: Common themes for signal transduction, cytoskeletal regulation, and membrane trafficking. *Annu. Rev. Cell. Dev. Biol.* **14**, 231–264 (1998).
4. Rameh, L. E. & Cantley, L. C. The role of phosphoinositide 3-kinase lipid products in cell function. *J. Biol. Chem.* **274**, 8347–8350 (1999).
5. Fruman, D. A., Meyers, R. E. & Cantley, L. C. Phosphoinositide kinases. *Annu. Rev. Biochem.* **67**, 481–507 (1998).
6. Fukui, Y., Ihara, S. & Nagata, S. Downstream of phosphatidylinositol-3 kinase, a multifunctional signaling molecule, and its regulation in cell responses. *J. Biochem. (Tokyo)* **124**, 1–7 (1998).
7. Vanhaesebroeck, B. *et al.* Synthesis and function of 3-phosphorylated inositol lipids. *Annu. Rev. Biochem.* **70**, 535–602 (2001).
8. Borggrefe, T., Wabl, M., Akhmedov, A. T. & Jessberger, R. A B-cell-specific DNA recombination complex. *J. Biol. Chem.* **273**, 17025–17035 (1998).
9. Borggrefe, T. *et al.* Cellular, intracellular, and developmental expression patterns of murine SWAP-70. *Eur. J. Immunol.* **29**, 1812–1822 (1999).
10. Masat, L. *et al.* Association of SWAP-70 with the B cell antigen receptor complex. *Proc. Natl Acad. Sci. USA* **97**, 2180–2184 (2000).
11. Borggrefe, T., Keshavarzi, S., Gross, B., Wabl, M. & Jessberger, R. Impaired IgE response in SWAP-70-deficient mice. *Eur. J. Immunol.* **8**, 2467–2475 (2001).
12. Shirai, R. *et al.* Synthesis of diacylglycerol analogs of phosphatidylinositol 3,4,5-trisphosphate. *Tetrahedron Lett.* **39**, 9485–9488 (1998).
13. Staal, S. Molecular cloning of the *akt* oncogene and its human homologues AKT1 and AKT2: amplification of AKT1 in a primary human gastric adenocarcinoma. *Proc. Natl Acad. Sci. USA* **84**, 5034–5037 (1987).
14. Tanaka, K. *et al.* A target of phosphatidylinositol 3,4,5-trisphosphate with a zinc finger motif similar to that of the ADP-ribosylation-factor GTPase-activating protein and two pleckstrin homology

domains. *Eur. J. Biochem.* **245**, 512–519 (1997).

15. Maekawa, M. *et al.* A novel mammalian Ras GTPase-activating protein which has phospholipid-binding and Btk homology regions. *Mol. Cell. Biol.* **14**, 6879–6885 (1994).
16. Campbell, D. & Kernan, J. Mast cells in the central nervous system. *Nature* **210**, 756–757 (1966).
17. Thomas, J. *et al.* Colocalization of X-linked agammaglobulinemia and X-linked immunodeficiency genes. *Science* **261**, 355–358 (1993).
18. Han, J. *et al.* Role of substrates and products of PI 3-kinase in regulating activation of Rac-related guanosine triphosphatases by Vav. *Science* **279**, 558–560 (1998).
19. Fleming, I. N., Gray, A. & Downes, C. P. Regulation of the Rac1-specific exchange factor Tiam1 involves both phosphoinositide 3-kinase-dependent and -independent components. *Biochem. J.* **351**, 173–182 (2000).
20. Das, B. *et al.* Control of intramolecular interactions between the pleckstrin homology and Dbl homology domains of Vav and Sos1 regulates Rac binding. *J. Biol. Chem.* **275**, 15074–15081 (2000).
21. Joly, M., Kazlasukas, A., Fay, F. S. & Corvera, S. Disruption of PDGF receptor trafficking by mutation of its PI-3 kinase binding sites. *Science* **263**, 684–687 (1994).
22. Martin, S. S. *et al.* Phosphatidylinositol 3-kinase is necessary and sufficient for insulin-stimulated stress fiber breakdown. *Endocrinology* **137**, 5045–5054 (1996).
23. Wennstrom, S. & Downward, J. Role of phosphoinositide 3-kinase in activation of ras and mitogen-activated protein kinase by epidermal growth factor. *Mol. Cell. Biol.* **19**, 4279–4288 (1999).
24. Missy, K. *et al.* Lipid products of phosphoinositide 3-kinase interact with Rac1 GTPase and stimulate G12P dissociation. *J. Biol. Chem.* **273**, 30279–30286 (1998).
25. Scita, G. *et al.* EPS8 and E3B1 transduce signals from Ras to Rac. *Nature* **401**, 290–293 (1999).
26. Iwamatsu, A. S-carboxymethylation of proteins transferred onto polyvinylidene difluoride membranes followed by *in situ* protease digestion and amino acid microsequencing. *Electrophoresis* **13**, 142–147 (1992).
27. Tanaka, K. *et al.* Evidence that a phosphatidylinositol 3,4,5-trisphosphate-binding protein can function in nucleus. *J. Biol. Chem.* **274**, 3919–3922 (1999).
28. Kita, Y. *et al.* Microinjection of activated phosphatidylinositol-3 kinase induces process outgrowth in rat PC12 cells through the Rac-JNK signal transduction pathway. *J. Cell Sci.* **111**, 907–915 (1998).
29. Lenzen, C., Cool, R. & Wittinghofer, A. Analysis of intrinsic and CDC25-stimulated guanine nucleotide exchange of p21ras-nucleotide complexes by fluorescence measurements. *Methods Enzymol.* **255**, 95–109 (1995).
30. Zheng, Y., Bagrodia, S. & Cerione, R. Activation of phosphoinositide 3-kinase activity by Cdc42Hs binding to p85. *J. Biol. Chem.* **269**, 18727–18730 (1994).
31. Gross, B. *et al.* SWAP-70 deficient mast cells are impaired in development and IgE-mediated degranulation. *Eur. J. Immunol.* (in the press).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank B. Mayer and M. Matsuda for useful suggestions and critical reading of the manuscript, and T. Nagase (Kazusa DNA Research Institute) for supplying the KIAA0640 clone. This work was supported by grants-in-aid for scientific research to Y.F. from the Ministry of Education, Science, Sports, and Culture of Japan, and by an NIH grant to R.J.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.F. (e-mail: ayfukui@mail.ecc.u-tokyo.ac.jp).

A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal

Y. Amy Lam⁺, T. Glen Lawson[†], Murugesan Velayutham[‡], Jay L. Zweier[‡] & Cecile M. Pickart^{*}

^{*} Department of Biochemistry and Molecular Biology, School of Public Health; and [‡] Department of Medicine and EPR Center, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

[†] Department of Chemistry, Bates College, Lewiston, Maine 04240, USA

The 26S proteasome is the chief site of regulatory protein turnover in eukaryotic cells¹. It comprises one 20S catalytic complex (composed of four stacked rings of seven members) and two axially positioned 19S regulatory complexes (each containing about 18 subunits) that control substrate access to the catalytic

chamber². In most cases, targeting to the 26S proteasome depends on tagging of the substrate with a specific type of polyubiquitin chain^{3–6}. Recognition of this signal is followed by substrate unfolding and translocation, which are presumably catalysed by one or more of six distinct AAA ATPases located in the base—a ring-like 19S subdomain that abuts the axial pore of the 20S complex and exhibits chaperone activity *in vitro*^{7–9}. Despite the importance of polyubiquitin chain recognition in proteasome function, the site of this signal's interaction with the 19S complex has not been identified previously. Here we use crosslinking to a reactive polyubiquitin chain to show that a specific ATPase subunit, S6' (also known as Rpt5), contacts the bound chain. The interaction of this signal with 26S proteasomes is modulated by ATP hydrolysis. Our results suggest that productive recognition of the proteolytic signal, as well as proteasome assembly and substrate unfolding, are ATP-dependent events.

We used intact 26S proteasomes—proteases with a relative molecular mass of 2,100,000 (M_r 2,100K)—as the starting point to identify, by means of crosslinking, the subunit(s) of the 19S complex that contacts the bound polyubiquitin chain. To construct a reactive version of tetra-ubiquitin (Ub_4), which is the minimal signal for efficient targeting to proteasomes⁶, we took advantage of the absence of endogenous cysteines in ubiquitin and our previous finding that a P37C mutation is permissive for chain recognition¹⁰. We placed the ubiquitin P37C mutant at the proximal position in Ub_4 and then introduced a radio-iodinated photoreactive cross-linker (APDP) at Cys 37 via disulphide exchange (Fig. 1). The reactive group was thus positioned near the proximal Leu 8 (refs 11, 12), whose side chain is necessary for optimal recognition of Ub_4 by 26S proteasomes^{6,13}.

Purified bovine 26S proteasomes were pre-incubated with a stoichiometric concentration (relative to the 19S complex) of the reactive chain, irradiated with ultraviolet radiation, and resolved by SDS–polyacrylamide gel electrophoresis (PAGE) under reducing

conditions to visualize crosslinked partner(s) at their normal molecular masses (Fig. 1). Despite the presence of approximately 18 distinct polypeptides in the 19S complex, there were only three significant crosslinked products, indicating substantial specificity of interaction (Fig. 2a, lane 6). The most abundant product was derived from intramolecular crosslinking, as it co-migrated with Ub_4 and was seen in the absence of proteasomes (Fig. 2a; * Ub_4 , lane 3 and 6). Reactions containing proteasomes and reactive Ub_4 uniquely yielded radiolabelled bands with molecular masses of roughly 48K and 50K (lane 6). The masses of these two proteins are consistent with identification as subunits of the 19S complex². Crosslinking of both proteins depended strictly on ultraviolet irradiation (not shown), and neither protein crosslinked to Ub_1 (lane 5), which does not bind detectably to proteasomes⁶. Reaction of the 50K protein, but not the smaller protein, was competed efficiently by unlabelled wild-type Ub_4 (Fig. 2a, compare lanes 8 and 6, and b), but not by a mutant Ub_4 that does not bind to proteasomes¹⁰ (Fig. 2a, compare lanes 9 and 6). Only the 50K

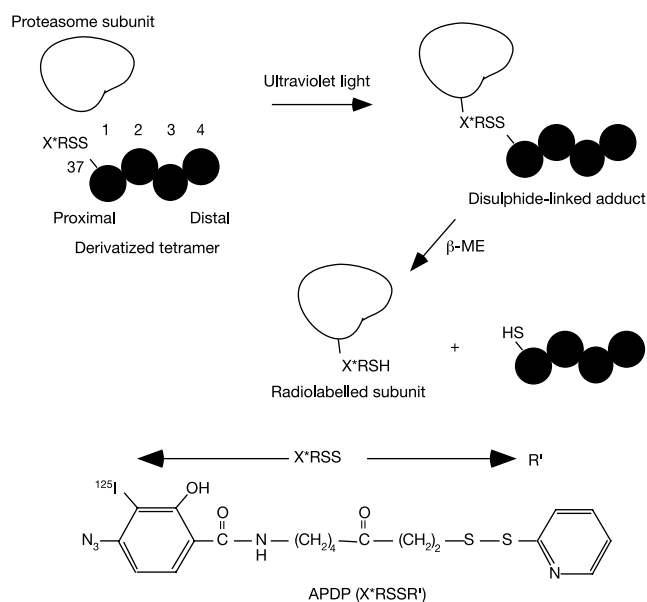


Figure 1 Crosslinking strategy. ¹²⁵I-labelled APDP was linked to Cys 37 in the proximal ubiquitin of Ub_4 by means of disulphide exchange (1 and 4 denote proximal and distal ubiquitins, respectively). As in the natural chain signal^{3,4}, the ubiquitins are linked by Lys 48–Gly 76 isopeptide bonds²⁷. After ultraviolet activation in the presence of proteasomes, the nitrene group of APDP crosslinks to a 19S subunit that is in close spatial proximity. Treatment of this initial crosslinked product with β -mercaptoethanol (β -ME) releases the (unlabelled) chain.

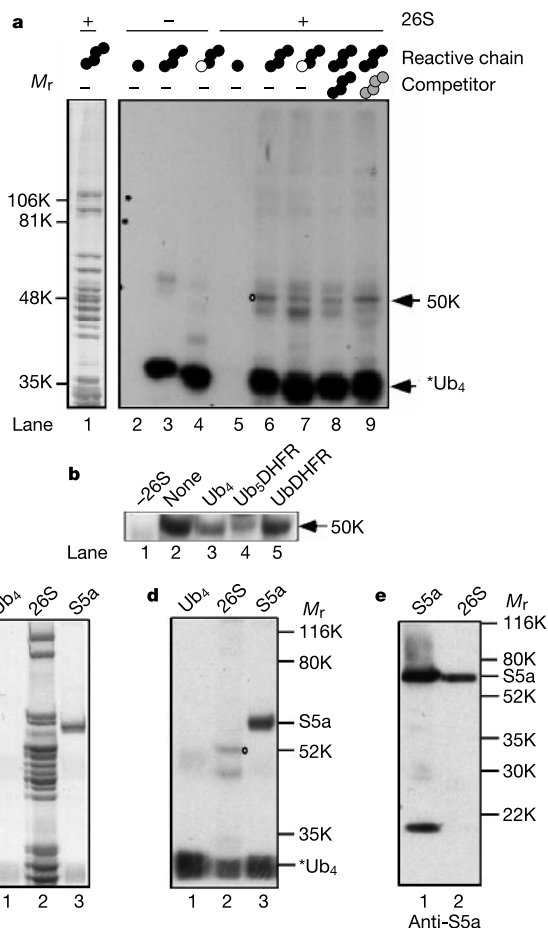


Figure 2 Specific crosslinking of Ub_4 to proteasomes. **a**, The 50K proteasome subunit crosslinks to Ub_4 . Crosslinking reactions (0.5 mM Mg-ATP) contained the indicated reactive species (Ub_1 , wild-type Ub_4 , or Ub_4 with proximal ubiquitin L8A (white circles)). Black circles, wild-type ubiquitin; grey circles, L8A,144A-ubiquitin. Lane 1, Coomassie stain of lane 6; lanes 2–9, autoradiograph. Lane 8 shows competition by unlabelled wild-type Ub_4 ; lane 9 shows competition by L8A, 144A Ub_4 (5 μ M each). The 50K band is indicated by a dot in the autoradiograph. * Ub_4 , product of intramolecular crosslinking (identical migration of this band on a non-reducing gel excluded its origin in an intermolecular reaction). **b**, The indicated competitors were added at 10 μ M each. **c**, Coomassie stain of Ub_4 crosslinking reactions with 1 μ M proteasomes or S5a (ref. 13) (2 mM Mg-ATP). Ub_4 denotes reaction with chain alone; S1, 2 and S3–12 denote subunits of the 19S complex. **d**, Autoradiograph of **c**. **e**, Anti-S5a western blot.

band showed diminished reaction when the L8A mutation was introduced into the proximal ubiquitin of the reactive chain (Fig. 2a, compare lanes 6 and 7)—an alteration that diminishes by several-fold the affinity of Ub₄ for proteasomes⁶. Of note, crosslinking by wild-type Ub₄ was competed efficiently by Ub₅-dihydrofolate reductase (DHFR), a polyubiquitinated substrate with high affinity for proteasomes⁶, but not by UbDHFR, which does not bind detectably⁶ (Fig. 2b). Thus, only the 50K protein behaved as expected for an authentic polyubiquitin recognition factor, whose crosslinking should be sensitive to chain length, competed by an authentic substrate, and dependent on a hydrophobic contact with ubiquitin Leu 8 (ref. 6). To exclude the possibility that the 50K subunit was adjacent to a primary binding protein and reacted only owing to the long linker arm of APDP, we showed that the 50K subunit still crosslinked when we moved APDP to the distal ubiquitin in the chain (not shown).

Purified S5a (also known as Rpn10), an approximately 50K subunit of the 19S complex, avidly binds polyubiquitin chains^{13,14}. Not surprisingly, purified recombinant S5a (Fig. 2c, lane 3) also

crosslinked strongly to the reactive chain (Fig. 2d). However, although S5a was present in the proteasome preparation (Fig. 2e), the crosslinked 50K protein in the 19S complex was not S5a, as shown by distinct mobilities of the respective crosslinked products (Fig. 2d, compare lanes 2 and 3). These results suggest that interactions between neighbours within the 19S complex occlude the polyubiquitin-binding site of S5a, and indicate that isolated S5a exhibits properties that are not manifested when this subunit is part of the 19S complex. The current results may explain the dispensability of the polyubiquitin-binding site of S5a subunits for the degradation of most polyubiquitinated substrates in yeast cells^{15,16}.

The 50K polyubiquitin-interacting protein in the 19S complex displayed an isoelectric point (pI) of roughly 5 on a two-dimensional IEF/SDS-PAGE gel (not shown). The only known 19S subunits with these physical properties are the ATPases S6' (also known as Rpt5 and TBP1) and S6 (also known as Rpt3 and TBP7)², which are thought to be located at adjacent positions in the presumptive ATPase ring^{17,18}. Peptide sequencing revealed that both S6 and S6' co-migrated with the labelled band (not shown). To determine which ATPase(s) was targeted by the reactive chain, we took advantage of the fact that the chain and its interaction partner are initially linked by a disulphide bond (Fig. 1). Thus, a fraction of about 80K in non-reducing SDS-PAGE, and this form should be immunochemically detectable using antibodies that distinguish the two ATPases (Fig. 3a, modulator complex lacks S6 (ref. 19). As shown in Fig. 3b, only the S6' antibody detected an 80K band in non-reducing SDS-PAGE of crosslinked proteasomes. Consistent with the identification of this band as a disulphide-linked adduct of S6' and Ub₄ it reacted with ubiquitin antibodies (Fig. 3c) and the principal fraction of the S6'/ubiquitin immunoreactivity disappeared after reduction (Fig. 3b, c, compare lanes 6 and 4). Therefore S6' ATPase within the 26S proteasome contacts

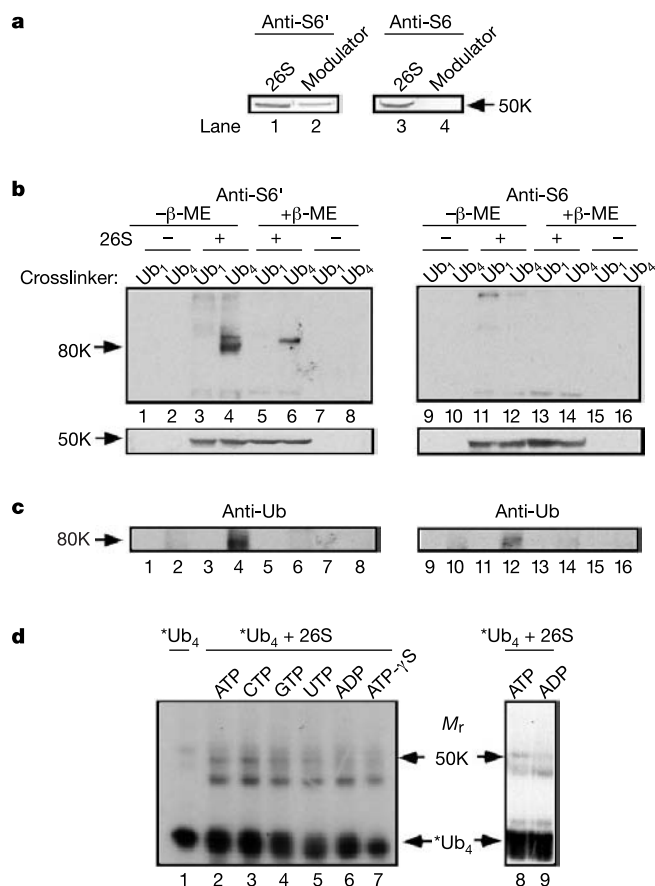


Figure 3 The 50K crosslinked protein is S6'. **a**, S6 and S6' antibodies do not crossreact. A western analysis of 26S proteasomes (lanes 1, 3) or equimolar modulator, a small complex that lacks S6 (ref. 19; lanes 2, 4), is shown. **b**, S6', but not S6, is initially linked to reactive Ub₄ by a disulphide bond (western blots). Crosslinking reactions with reactive Ub₄ or Ub₄ (2 mM Mg-ATP) were quenched in SDS with or without β-mercaptoethanol (β-ME). Sections of the blot corresponding to greater than 60K (top panel) and about 50K (bottom panel) were analysed separately. **c**, The 80K complex contains ubiquitin (anti-ubiquitin western blots; the top panels from **b** were stripped and re-probed). **d**, Efficient crosslinking requires ATP hydrolysis (autoradiograph). 26S proteasomes were depleted of ATP before crosslinking with reactive Ub₄ plus the indicated nucleotide (2 mM, with 10 mM MgCl₂). Lanes 8 and 9 show an independent experiment. *Ub₄, product of intramolecular crosslinking.

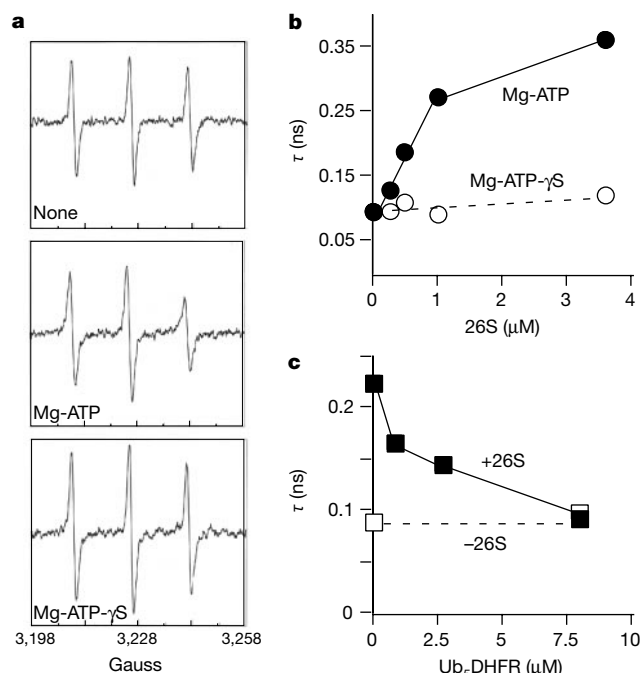


Figure 4 ATP hydrolysis modulates the polyubiquitin chain-proteasome interaction (EPR). **a**, Representative spectra. Incubations contained 2 μM Ub₄-proxyl; middle and lower panels also contained 3.6 μM 26S proteasomes with 1 mM Mg-ATP or ATP-γS (plus ATP regenerating system). **b**, Titration analysis. Incubations contained 2 μM Ub₄-proxyl, proteasomes and Mg-ATP or Mg-ATP-γS. **c**, Competition study. Incubations contained 1.7 μM Ub₅-proxyl, 2.7 μM proteasomes and Ub₅DHFR⁶ as shown.

the bound polyubiquitin chain in a specific manner (Figs 2a, b and 3). However, Ub₄ did not bind detectably to recombinant glutathione S-transferase (GST)–Rpt5 in a pull-down assay (not shown), suggesting that the ability to form a functional chain–interaction site on S6' is modulated by the presence of this subunit's neighbours within the 19S complex.

Consistent with identification of the S6' ATPase as the subunit of the 19S complex that interacts with the polyubiquitin chain, efficient crosslinking to S6' required ATP hydrolysis, as indicated by diminished crosslinking of the 50K protein in the presence of ADP or ATP-γS (Fig. 3d). A series of nucleoside 5'-triphosphates (NTPs) supported crosslinking in the order ATP ≈ CTP > GTP > UTP (Fig. 3d), similar to the NTPase specificity of the 19S complex²⁰. ATP-γS is able to maintain the 19S–20S association (our own unpublished data and ref. 21); therefore, the diminished crosslinking signal seen in the presence of APT-γS (Fig. 3d) is not due to impaired proteasomal integrity. ATP hydrolysis may rather be necessary for efficient binding of the polyubiquitin signal to its cognate site. To address this possibility, we introduced a spin label (proxyl nitroxide coupled to maleimide) into Ub₄ and monitored its electron paramagnetic resonance (EPR) spectrum in the presence of proteasomes. In the absence of proteasomes, Ub₄-proxyl displayed a spectrum typical of a rapidly tumbling molecule (Fig. 4a, top panel). Upon addition of 26S proteasomes in the presence of Mg-ATP, the spectrum became more anisotropic, as demonstrated by the decreased ratio of the high field peak to the central peak, and the increased rotational correlation time (τ) (Fig. 4a, middle panel, and b). This effect was eliminated when ATP was replaced with ATP-γS (Fig. 4a, bottom panel, and b), providing direct physical evidence that the chain–proteasome interaction is modulated by ATP hydrolysis. To ensure that the chain binding site monitored in this experiment was functionally significant, we showed that the Mg-ATP-dependent increase in τ was competed efficiently by Ub₃DHFR (Fig. 4c), indicating that the chain–interaction site monitored by EPR is that which binds Ub₃DHFR with high affinity⁶. We note that even in the presence of proteasomes and Mg-ATP, the EPR signal of the spin-labelled chain is not characteristic of a highly immobilized molecule²², suggesting that free and proteasome-bound chains are in rapid equilibrium. Consistent with this model, Ub₄ did not co-sediment with 26S proteasomes on a glycerol gradient (our own unpublished data).

We have shown that a polyubiquitin chain interacts with the base of the 19S complex in a manner that places a critical region of the chain in close proximity to the S6' ATPase (Figs 2 and 3). Moreover, ATP hydrolysis modulates the interaction between the chain and the proteasome (Figs 3d and 4). Competition by the 26S proteasome substrate Ub₃DHFR in crosslinking and EPR provides evidence that the site being monitored is that responsible for the recognition of polyubiquitinated substrates. These results implicate S6' specifically, and the base of the 19S complex generally, in proteolytic signal recognition. S6' is thus one of only a handful of 19S subunits with assigned functions. A specific role for this subunit in ATP-modulated polyubiquitin recognition is consistent with the ability of a non-proteasomal AAA ATPase to bind polyubiquitin chains²³, and may also help to explain the severe inhibition of model substrate degradation caused by a non-conservative mutation in the ATP-binding site of yeast Rpt5 (ref. 24). (However, because the ATPases probably form a ring, the effect of ATP on chain interaction could originate in the ATP site of a different ATPase.) Binding of the polyubiquitin signal to the base of the 19S complex should efficiently position the substrate for downstream events in proteolysis, as this subdomain also mediates substrate unfolding^{7,9} and gating of the axial pore of the 20S complex²⁵. A requirement for the lid of the 19S complex in the degradation of polyubiquitin-tagged substrates⁸ may reflect as yet undiscovered interactions between a lid subunit(s) and the chain or substrate. Interestingly, the homogeneous ATPase subunits of the simpler archaeobacterial and eubacterial proteasomes

also mediate substrate recognition, usually by binding to a short motif within the substrate polypeptide chain². Our results not only document a new role for ATP in polyubiquitin-dependent proteolysis, but also suggest that delegation of a signal-recognition function to a specific ATPase accompanied the functional diversification of these subunits in eukaryotes^{24,25}. □

Methods

Antibodies and proteins

We purified 26S proteasomes from bovine erythrocytes²⁶. Antibodies against S6 and S6' were from Affiniti. We produced the antibodies against ubiquitin. Antibodies against Rpn10/S5a were a gift of R. Vierstra.

Reactive chains

Polyubiquitin chains were synthesized enzymatically^{6,27}. APDP (N-[4-(p-azido-salicylamido)butyl]-3'-(2'-pyridyldithio)propionamide, 20 mM in DMSO; Pierce) was diluted to 3.4 mM in 35 μl of 0.15 M phosphate (pH 7.4) and radio-iodinated using chloramine-T (0.3 mCi of Na¹²⁵I for 5 min, followed by 5 mM unlabelled KI). Ub₄ (proximal ubiquitin P37C, 4 mg ml⁻¹) was reduced with 0.5 mM dithiothreitol (30 min, 37 °C), diluted to 50 μM, incubated with ¹²⁵I-labelled APDP (2 mM, 1 h, 37 °C), and separated from unincorporated APDP on a spin column to yield APDP-Ub₄ at approximately 10⁶ c.p.m. μg⁻¹. We used similar procedures to make reactive Ub₁. Reactive molecules were stored in the dark at 5 °C and used within 1 month.

Crosslinking

26S proteasomes (0.5 μM) were pre-incubated in the dark with reactive Ub₁ or Ub₄ (1 μM) for 10 min (10 μl, pH 7.6, 37 °C) in the presence of 0.5–2 mM Mg-ATP plus 1 μM ubiquitin aldehyde. To induce crosslinking, tubes were placed on an ultraviolet transilluminator for 2.5 min. Reactions were immediately quenched with SDS sample buffer (with or without mercaptoethanol as indicated) before resolution by SDS-PAGE.

EPR measurements

Maleimido-proxyl nitroxide (Aldrich) was reacted with Cys 37 (proximal) in Ub₄ and unincorporated spin label was removed by dialysis. EPR samples (5 μl) were loaded into capillary tubes with an internal diameter of 0.4 mm, and measurements were performed on a Bruker 300E spectrometer with loop-gap resonator at X-band using 1 mW microwave power²⁸. In Fig. 4 spectra consisted of 25 or 50 summed 30-s scans. Each point in Fig. 4b was obtained by averaging 75 scans (from two experiments). In Fig. 4c, spectra consisted of ten scans (to prevent substantial degradation of Ub₃DHFR). We calculated rotational correlation times, τ , as described²².

Received 17 December 2001; accepted 13 February 2002.

- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Voges, D., Zwickl, P. & Baumeister, W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68**, 1015–1068 (1999).
- Chau, V. et al. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576–1583 (1989).
- Finley, D. et al. Inhibition of proteolysis and cell cycle progression in a multiubiquitination-deficient yeast mutant. *Mol. Cell. Biol.* **14**, 5501–5509 (1994).
- Pickart, C. M. Ubiquitin in chains. *Trends Biochem. Sci.* **25**, 544–548 (2000).
- Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**, 94–102 (2000).
- Braun, B. C. et al. The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nature Cell Biol.* **1**, 221–226 (1999).
- Glickman, M. H. et al. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**, 615–623 (1998).
- Strickland, E., Hakala, K., Thomas, P. J. & DeMartino, G. N. Recognition of misfolded proteins by PA700, the regulatory subcomplex of the 26 S proteasome. *J. Biol. Chem.* **275**, 5565–5572 (2000).
- Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M. & Pickart, C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl. Acad. Sci. USA* **93**, 861–866 (1996).
- Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).
- Cook, W. J., Jeffrey, L. C., Kasperek, E. & Pickart, C. M. Structure of tetraubiquitin shows how multiubiquitin chains can be formed. *J. Mol. Biol.* **236**, 601–609 (1994).
- Beal, R. E., Toscano-Cantaffa, D., Young, P., Rechsteiner, M. & Pickart, C. M. The hydrophobic effect contributes to polyubiquitin chain recognition. *Biochemistry* **37**, 2925–2934 (1998).
- Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **269**, 7059–7061 (1994).
- Fu, H. et al. Multiubiquitin chain binding and protein degradation are mediated by distinct domains within the 26 S proteasome subunit Mcl1. *J. Biol. Chem.* **273**, 1970–1981 (1998).
- van Nocker, S. et al. The multiubiquitin-chain-binding protein Mcl1 is a component of the 26S proteasome in *Saccharomyces cerevisiae* and plays a nonessential, substrate-specific role in protein turnover. *Mol. Cell. Biol.* **16**, 6020–6028 (1996).
- Hartmann-Petersen, R., Tanaka, K. & Hendil, K. B. Quarternary structure of the ATPase complex of human 26S proteasomes determined by chemical cross-linking. *Arch. Biochem. Biophys.* **386**, 89–94 (2001).
- Davy, A. et al. A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep.* **2**, 821–828 (2001).

19. DeMartino, G. N. *et al.* Identification, purification, and characterization of a PA700-dependent activator of the proteasome. *J. Biol. Chem.* **271**, 3112–3118 (1996).
20. Hoffman, L. & Rechsteiner, M. Nucleotidase activities of the 26S proteasome and its regulatory complex. *J. Biol. Chem.* **271**, 32538–32545 (1997).
21. Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3429 (2000).
22. Qin, P. Z., Butcher, S. E., Feigon, J. & Hubbell, W. L. Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: a site-directed spin labeling study. *Biochemistry* **40**, 6929–6936 (2001).
23. Dai, R. M. & Li, C.-C. H. Valosin-containing protein is a multi-ubiquitin chain-targeting factor required in ubiquitin-proteasome degradation. *Nature Cell Biol.* **3**, 740–744 (2001).
24. Rubin, C. M., Glickman, M. H., Larsen, C. N., Dhruvakumar, S. & Finley, D. Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome. *EMBO J.* **17**, 4909–4919 (1998).
25. Kohler, A. *et al.* The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Mol. Cell* **7**, 1143–1152 (2001).
26. Hoffman, L., Pratt, G. & Rechsteiner, M. Multiple forms of the 20S multicatalytic and the 26S ubiquitin/ATP-dependent proteases from rabbit reticulocyte lysate. *J. Biol. Chem.* **267**, 22362–22368 (1992).
27. Piotrowski, J. *et al.* Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J. Biol. Chem.* **272**, 23712–23721 (1997).
28. Sankarapandi, S. & Zweier, J. L. Bicarbonate is required for the peroxidase function of Cu,Zn-superoxide dismutase at physiological pH. *J. Biol. Chem.* **274**, 1226–1232 (1999).

Acknowledgements

We thank G. DeMartino for a gift of modulator complex; A. Varshavsky for the GST-Rpt5 construct; A. Mildvan for discussions; and M. Hochstrasser and B. Cohen for comments on the manuscript. This work was supported by a grant from the National Institutes of Health. Y.A.L. is a senior fellow of the Leukemia and Lymphoma Society.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.P. (e-mail: cpickart@jhmi.edu).

Geobacter metallireducens accesses insoluble Fe(III) oxide by chemotaxis

Susan E. Childers, Stacy Ciufo & Derek R. Lovley

Department of Microbiology, University of Massachusetts, Amherst, Massachusetts 01003, USA

Microorganisms that use insoluble Fe(III) oxide as an electron acceptor can have an important function in the carbon and nutrient cycles of aquatic sediments and in the bioremediation of organic and metal contaminants in groundwater^{1,2}. Although Fe(III) oxides are often abundant, Fe(III)-reducing microbes are faced with the problem of how to access effectively an electron acceptor that can not diffuse to the cell. Fe(III)-reducing microorganisms in the genus *Shewanella* have resolved this problem by releasing soluble quinones that can carry electrons from the cell surface to Fe(III) oxide that is at a distance from the cell^{3,4}. Here we report that another Fe(III)-reducer, *Geobacter metallireducens*, has an alternative strategy for accessing Fe(III) oxides. *Geobacter metallireducens* specifically expresses flagella and pili only when grown on insoluble Fe(III) or Mn(IV) oxide, and is chemotactic towards Fe(II) and Mn(II) under these conditions. These results suggest that *G. metallireducens* senses when soluble electron acceptors are depleted and then synthesizes the appropriate appendages to permit it to search for, and establish contact with, insoluble Fe(III) or Mn(IV) oxide. This approach to the use of an insoluble electron acceptor may explain why *Geobacter* species predominate over other Fe(III) oxide-reducing micro-

organisms in a wide variety of sedimentary environments^{5–8}.

Geobacter metallireducens, the first organism found to completely oxidize organic compounds to carbon dioxide with Fe(III) oxide serving as the electron acceptor^{9–11}, was previously reported to be non-motile¹². However, in those previous studies, motility was evaluated in cultures grown on soluble electron acceptors such as nitrate or Fe(III)-citrate. We observed with phase-contrast microscopy that cells of *G. metallireducens* that had been grown on insoluble Fe(III) were motile. Furthermore, when we placed cells that had been grown on Fe(III) or Mn(IV) oxide onto soft agar plates they swam out from the line on which they were placed, whereas cells that had been grown on soluble Fe(III) did not (Fig. 1a).

Cells grown with insoluble Fe(III) or Mn(IV) oxides as the electron acceptor had lateral flagella that were apparent with transmission electron microscopy, whereas cells grown with soluble Fe(III) lacked flagella (Fig. 1b). Examination of 100 cells from each type of culture revealed that 77% and 68% of cells grown with

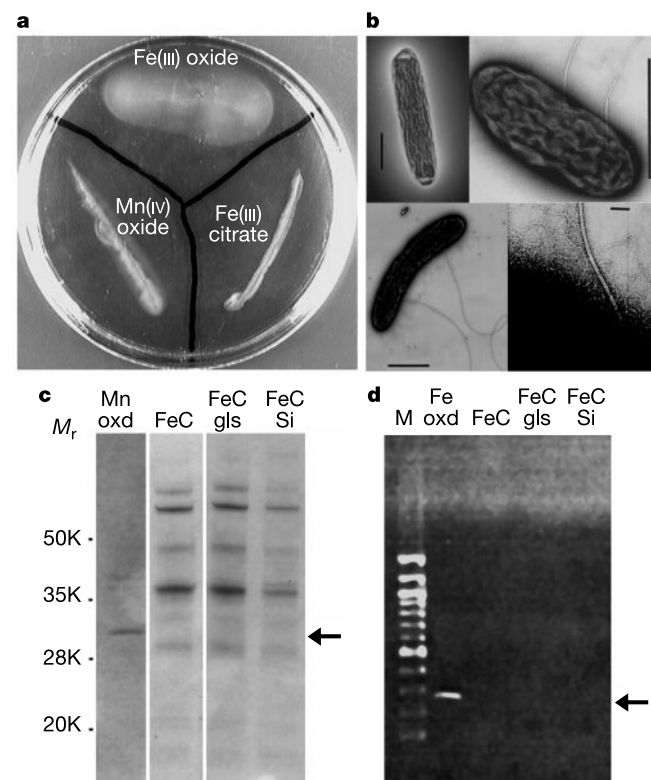


Figure 1 Expression of flagella and pili by *G. metallireducens*. **a**, Motility agar plate of swimming motility in response to growth with the electron acceptor indicated. A heavy suspension of cells was placed onto the plate and, after several days, the appearance of a light-grey haze extending outward from the initial streak indicated swimming. **b**, Electron micrographs showing the absence of flagella on cells grown with Fe(III)-citrate (top left), in contrast to cells grown with Fe(III) (top right) or Mn(IV) (bottom left) oxides as the terminal electron acceptor. Scale bars, 1 μ m. The bottom right panel is a higher resolution electron micrograph of pili on cells. Scale bar, 0.1 μ m. The contrast of the image was increased to enhance the visibility of pili. Cells were stained with 4% uranyl acetate and viewed with a JEOL 100S microscope. **c**, SDS-polyacrylamide gel electrophoresis of surface proteins obtained from *G. metallireducens* cells grown on Mn(IV) oxide (Mn oxd), Fe(III)-citrate (FeC), and Fe(III) citrate containing glass beads (FeC gls) or silica (FeC Si). The arrow points to the band consisting of flagellin protein. The molecular size standards are indicated on the left. **d**, PCR amplification of *pilA* from cDNA generated from mRNA. Cells were grown with the electron acceptor indicated (labelling as in **c**). Controls in which RNase was added to the reactions before reverse transcription was carried out showed no bands, indicating that the RNA preparation was not contaminated with DNA (not shown). The arrow points to the *pilA* PCR product. The marker (M) is a 1-kb DNA ladder (New England Biolabs).

insoluble Fe(III) or Mn(IV) oxides as the electron acceptor had flagella; the actual percentage of flagellated cells is higher because flagella are easily sheared off during preparation of cells for electron microscopy. In contrast, 90% of the cells grown with soluble Fe(III)-citrate lacked flagella. The cells grown on Mn(IV) (Fig. 1c) and Fe(III) (data not shown) oxides had surface proteins with a relative molecular mass of 31,000 (M_r , 31K), the size expected for the flagellin subunit based on the flagellin homologue identified in the genome sequence of the close relative, *Geobacter sulfurreducens* (<http://www.tigr.org>). Mass spectrometry confirmed that the protein was flagellin. This protein was not present in surface proteins of cells grown on soluble Fe(III) nor cells grown on soluble Fe(III) in a medium that also contained glass beads or silica, which were included to mimic the solid surface provided by the oxides (Fig. 1c).

The Fe(III) and Mn(IV) oxide-grown cultures also had pili along one side of the cells (Fig. 1b). Approximately 66% and 83%, respectively, of Fe(III) and Mn(IV) oxide-grown cells were piliated, whereas no pili were evident on cells grown on soluble Fe(III), with or without added glass beads or silica. The genome of the closely related bacterium *G. sulfurreducens* was found to contain genes for the biosynthesis of type IV pili, appendages known to have a function in attachment and movement along surfaces in other microorganisms^{13–15}. Using sequences obtained from the *G. sulfurreducens* genome, polymerase chain reaction (PCR) primers were designed to amplify the putative *pilA* gene that codes for pilin, the pilus subunit¹⁶, from *G. metallireducens*. PCR amplification of DNA from both *G. sulfurreducens* and *G. metallireducens* yielded one product of the expected size (approximately 400 base pairs) with the same sequence. Messenger RNA for *pilA* was detected in cells grown with Fe(III) (Fig. 1d) or Mn(IV) (not shown) oxides, but not soluble Fe(III) (Fig. 1d). Fe(II) was not the signal for production of flagella and pili, as Fe(II) was abundant in the Fe(III)-citrate cultures and was not present in the cultures grown on Mn(IV). These results suggest further that the formation of pili and flagella is regulated in response to the electron acceptor that is available.

The specific production of appendages associated with motility only when insoluble Fe(III) or Mn(IV) oxides served as the electron acceptor suggested that these appendages might aid *G. metallireducens* in searching for, and establishing contact with, Fe(III) or Mn(IV) oxides. Although Fe(III) and Mn(IV) oxides are highly insoluble, as they are reduced they release Fe(II) and Mn(II),

which are soluble. Thus, it is expected that there will be a gradient of soluble Fe(II) and Mn(II) emanating from Fe(III) and Mn(IV) oxides under anoxic conditions. In chemotaxis assays, motile cells accumulated around agarose plugs containing soluble Fe(II) and Mn(II), but not insoluble Fe(III) or Mn(IV) (Fig. 2). Cells were not attracted to soluble Fe(III) in the form of Fe(III)-citrate (Fig. 2) or Fe(III)-nitrilotriacetic acid, nor to nitrate, an electron acceptor used preferentially over Fe(III) (not shown). To our knowledge, this is the first demonstration that microorganisms may be chemotactic towards Fe(II) or Mn(II).

Other Fe(III) oxide-reducing microorganisms such as *Shewanella* species^{3,17} and *Geothrix fermentans*⁴ overcome the insolubility of Fe(III) oxide by releasing soluble electron-shuttling compounds that can transfer electrons from the cell surface to Fe(III) oxides located at a significant distance from the cell. *Shewanella* species may be motile when grown under aerobic conditions and chemotactic towards the soluble electron acceptors nitrate and nitrite¹⁸. However, we observed both microscopically and with motility plate assays that neither *Shewanella* nor *G. fermentans* are motile when grown with Fe(III) oxide as the electron acceptor. This is consistent with the concept that these organisms can reduce Fe(III) oxide by means of an electron shuttle.

An electron shuttle represents an energetic cost in terms of the organic carbon devoted to shuttle synthesis that might otherwise be oxidized in energy-yielding catabolism. This energetic investment may be recouped if the electron-shuttling compound can be recycled through successive reduction and oxidation cycles. In subsurface environments, however, electron-shuttling compounds are likely to be lost from the point of release by diffusion and advection, as well as by nonspecific adsorption onto soil particles and degradation by other microorganisms. Thus, releasing an electron-shuttling compound in subsurface environments could easily result in a net energetic loss. In comparison, the energy requirement for flagellar synthesis and motility is generally regarded to be a small percentage of the overall energetic cost of cell synthesis and maintenance¹⁹. The finding that *Geobacter* species predominate over *Shewanella* and *Geothrix* species in a wide variety of subsurface environments^{5–8}, suggests that, in such environments, accessing Fe(III) oxides through chemotaxis and attachment may be a more adaptive strategy for reducing insoluble Fe(III) than transferring electrons with an electron shuttle. □

Methods

Culture methods

We grew *G. metallireducens* anaerobically in fresh water acetate (FWA) medium as described¹⁰. Motility plates were composed of FWA medium containing 50 mM Fe(III)-citrate and 0.35% agar. All plate manipulations were done in a 30 °C heated anaerobic glove box (Coy) containing a N₂:CO₂:H₂ (in per cent, 83:10:7) atmosphere.

Expression analyses

Surface proteins were prepared as described previously²⁰. Cells were collected by centrifugation, re-suspended in 1–2 ml 10 mM Tris, pH 7.4, and vortexed for 2 min. Cells were removed by centrifugation at 16,000g for 1 min. The supernatant was transferred to a new tube and centrifugation was repeated for 5 min. The resulting cell-free supernatant was transferred to a new tube and MgCl₂ was added to 100 mM. Tubes were kept at 4 °C overnight and precipitated proteins were collected by centrifugation at 16,000g for 15 min at 4 °C.

We used primers for *pilA* (forward primer, 5'-GGATCCATGCTGAACGTTT-3'; reverse primer, 5'-GGGATCCTCACCAGTCTTGTACGGAG-3') to amplify the gene from genomic DNA of both *G. sulfurreducens* and *G. metallireducens*. Cells were collected from cultures of *G. metallireducens* grown on various electron acceptors, and mRNA was extracted using an RNeasy kit (Qiagen). In the case of Fe(III) oxide, cultures were first treated with oxalate to solubilize the insoluble Fe(III) oxide particles²¹. Isolated mRNA was reverse transcribed with OmniScript (Qiagen), and the resulting complementary DNA was subjected to PCR amplification using primers internal to *pilA* (forward primer, 5'-GGC AATTGCCGTCCCAACT-3'; reverse primer, 5'-TTACCCTTGATCTCCTCGGT-3').

Chemotaxis assays

Chemotaxis assays were performed using a modified agarose plug method²². The sides of a chemotaxis chamber consisted of two plastic strips 15-mm apart on a glass slide. A drop of molten low-melting-point agarose containing the substrate to be tested was placed on a coverslip that was inverted onto the plastic strips to form a chamber. Cells for chemotaxis

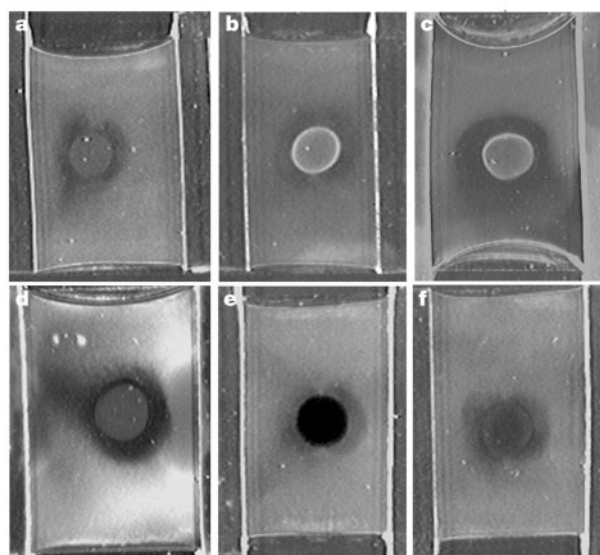


Figure 2 Chemotaxis to Fe(II) and Mn(II). Agarose plugs contained: **a**, chemotaxis buffer; **b**, 10 mM MnCl₂; **c**, 10 mM FeSO₄; **d**, 20 mM Fe(III) citrate; **e**, Mn(IV) oxide; **f**, Fe(III) oxide. Bright ring of cells around agarose plugs in **b** and **c** is indicative of chemotaxis.

assays were grown on Mn(IV) oxide, collected anaerobically, re-suspended in chemotaxis buffer, and supplied with 0.5 mM acetate as an energy source before being flooded into the chamber.

Received 20 December 2001; accepted 21 February 2002.

1. Thamdrup, B. Bacterial manganese and iron reduction in aquatic sediments. *Adv. Microb. Ecol.* **16**, 41–84 (2000).
2. Lovley, D. R. *Fe(III) and Mn(IV) Reduction in Environmental Metal–Microbe Interactions* (ed. Lovley, D. R.) 3–30 (ASM, Washington DC, 2000).
3. Newman, D. K. & Kolter, R. A role for excreted quinones in extracellular electron transfer. *Nature* **405**, 94–97 (2000).
4. Nevin, K. P. & Lovley, D. R. in *ASM 101st General Meeting* 588 (ASM, Washington DC, 2001).
5. Rooney-Varga, J. N., Anderson, R. T., Fraga, J. L., Ringelberg, D. & Lovley, D. R. Microbial communities associated with anaerobic benzene degradation in a petroleum-contaminated aquifer. *Appl. Microbiol.* **65**, 3056–3063 (1999).
6. Snoeyenbos-West, O. L., Nevin, K. P., Anderson, R. T. & Lovley, D. R. Enrichment of *Geobacter* species in response to stimulation of Fe(III) reduction in sandy aquifer sediments. *Microb. Ecol.* **39**, 153–167 (2000).
7. Rölting, W. F. M., van Breukelen, B. M., Braster, M., Lin, B. & van Verseveld, H. W. Relationships between microbial community structure and hydrochemistry in a landfill leachate-polluted aquifer. *Appl. Environ. Microbiol.* **67**, 4619–4629 (2001).
8. Stein, L. Y., La Duc, M. T., Grundl, T. J. & Neelson, K. H. Bacterial and archaeal populations associated with freshwater ferromanganous micronodules and sediments. *Environ. Microbiol.* **3**, 10–18 (2001).
9. Lovley, D. R., Stolz, J. F., Nord, G. L. Jr & Phillips, E. J. P. Anaerobic production of magnetite by a dissimilatory iron-reducing bacterium. *Nature* **330**, 252–254 (1987).
10. Lovley, D. R. & Phillips, E. J. P. P. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl. Environ. Microbiol.* **54**, 1472–1480 (1988).
11. Lovley, D. R. *et al.* Oxidation of aromatic contaminants coupled to microbial iron reduction. *Nature* **339**, 297–299 (1989).
12. Lovley, D. R. *et al.* *Geobacter metallireducens* gen. nov. sp. nov., a microorganism capable of coupling the complete oxidation of organic compounds to the reduction of iron and other metals. *Arch. Microbiol.* **159**, 336–344 (1993).
13. Henriksen, J. Twitching motility. *Annu. Rev. Microbiol.* **37**, 81–93 (1983).

14. Strom, M. S. & Lory, S. Structure-function and biogenesis of the type IV pili. *Annu. Rev. Microbiol.* **47**, 565–596 (1993).
15. Wall, D. & Kaiser, D. Type IV pili and cell motility. *Mol. Microbiol.* **32**, 1–10 (1999).
16. Alm, R. A. & Mattick, J. S. Genes involved in the biogenesis and function of type-4 fimbriae in *Pseudomonas aeruginosa*. *Gene* **192**, 89–98 (1997).
17. Nevin, K. P. & Lovley, D. R. in *ASM 100th General Meeting* 598 (ASM, Washington DC, 2000).
18. Neelson, K. H., Moser, D. P. & Saffarini, D. A. Anaerobic electron acceptor chemotaxis in *Shewanella putrefaciens*. *Appl. Environ. Microbiol.* **61**, 1551–1554 (1995).
19. Macnab, R. M. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* Vol. 1 (ed. Neidhardt, F. C. *et al.*) 123–145 (ASM, Washington DC, 1996).
20. Wu, S. S. & Kaiser, D. Regulation of expression of the *pilA* gene in *Myxococcus xanthus*. *J. Bacteriol.* **179**, 7748–7758 (1997).
21. Phillips, E. J. P. & Lovley, D. R. Determination of ferric and ferrous iron in oxalate extracts of sediment. *Soil Sci. Soc. Am. J.* **51**, 938–941 (1987).
22. Parales, R. E., Ditty, J. L. & Harwood, C. S. Toluene-degrading bacteria are chemotactic towards the environmental pollutants benzene, toluene and trichloroethylene. *Appl. Environ. Microbiol.* **66**, 4098–4104 (2000).

Acknowledgements

We thank L. Yin and the Central Microscopy Facility at the University of Massachusetts for assistance with transmission electron microscopy. Preliminary sequence data was obtained from The Institute for Genomic Research website at <http://www.tigr.org>. This research was supported with grants from the Office of Biological and Environmental Research of the Department of Energy and the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R.L. (e-mail: dlovley@microbio.umass.edu).

Time for analysis

Magnesium, oxygen and hydrogen are elements in this week's selection.

TOC-V analyser

Shimadzu www.shimadzu.com

Integrated sample pre-treatment

Shimadzu's TOC-V series laboratory TOC (total organic carbon) analysers incorporate an integrated sample pre-treatment system with an eight-port valve and a syringe with sparging capacity. Acidification and sparging take place within the instrument, eliminating the need for dilution of acid or alkaline samples and reducing the potential for user contact with concentrated acids. With auto-dilution, the instrument is able to measure from 4 µg to 25,000 mg per litre. It also performs up to 10-point calibrations using a standard TOC solution. It is available as a stand-alone or as a PC-controlled instrument.

SuperSignal III

Pierce www.piercenet.com

Speedy, non-radioactive RPA

Pierce has collaborated with Ambion to produce this non-isotopic ribonuclease protection assay (RPA) kit. Exposures can be made in minutes rather than overnight and probes remain stable for years, rather than weeks, with no need to run probe purifications. The assay uses a single-tube reaction that requires no phenol extraction or proteinase K digestions. A kit instruction book can be downloaded from Pierce's website.

Micro pH electrode

Lazar Research Laboratories

www.lazarlab.com

Small wonder

This electrode from Lazar Research Laboratories can measure samples as small as 5 µl — less than a tenth the volume of a single drop of liquid — as well as small biological specimens down to less than 0.5 mm in length. The electrode fits directly in 96-well plates,

microcentrifuge tubes, serum cups, capillary tubes and NMR tubes. Biological specimens that can be measured include various animal and plant tissues. Surface pH measurements on biological and chemical samples can also be made. The probe is made of sturdy fluoropolymer and has a 1 mm tip with a 2 mm diameter body. It can be used with any standard pH meter or strip chart recorder.

Finnpipette Focus

ThermoLabsystems www.thermo.com

Take the strain out of pipetting

The ergonomically designed Focus pipette features a short tip cone, a removable snap-on handle and a quick-click volume setting. Its short tip cone improves comfort by minimizing hand and arm strain during pipetting. The handle comes in three sizes, allowing the user to choose which one best fits his or her hand. Volume setting is fast and provides fine



Focus on the rack: all the sizes, in all the colours.

control over small increments. The pipette has a soft-touch tip ejector, which ejects the tip with a light push. The pipette is available in 10 volume ranges.

LCpro

ADC Bioscientific www.adc.co.uk

Travel light

ADC Bioscientific describes the LCpro as an intelligent, portable photosynthesis measurement system. The unit weighs only 4 kg and provides automatic and independent control of the environmental conditions within the plant leaf chamber, including CO₂, H₂O, temperature and PAR. It will function for up to 16 hours from a single charge, allowing photosynthesis experiments to be performed in many field sites. A range of interchangeable leaf chambers is available.

OSR6189

Olympus Diagnostic Systems

www.olympus.com

Improved magnesium assay

OSR6189 is a new magnesium reagent developed for use on all Olympus Diagnostics analysers. The reformulated assay uses stable, ready-to-use liquid reagents delivered in bar-coded vials for automatic information transfer. The new reagent has a linear range of 0.2–3.3 mmol per litre for serum, increasing the upper limit from the former 2.88 mmol per litre. Decreased blood and tissue levels of magnesium have been associated with high blood pressure, kidney stones, heart disease and heart attacks due to coronary artery spasm. Toxicity, although rare, may cause depression of the central nervous system resulting in symptoms which, in extreme cases, may prove fatal.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Ben Corp (4974)
Andy Douglas (4975)
Laura Pearson (4977)
Netherlands/ Italy/ Iberia:
Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatanmachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Collaborative conundrums

For today's researchers, the word collaboration is a loaded term. On the one hand, who could possibly be against it? But at the same time, as the climate within science moves towards larger projects that require shared resources, professional development and funding opportunities are increasingly tied to how well researchers work with others.

The structure of these collaborations and how they begin life might offer clues to their chance for success — or at least the level of involvement for the participating groups. In the 'bottom-up' approach, investigators agree jointly to apply for funds and then negotiate their own alliance. Alternatively, a granting body can set the parameters for cooperation in a 'top-down' structure.

In the United States, the chance of a windfall in biodefence money is creating some bottom-up collaborations (see pages 4–5). Researchers are realizing that by combining their expertise with others, they might stand a better chance of winning a share of this money — and be better placed to succeed once they are funded.

In Europe, the European Commission (EC), in its proposal for the sixth Framework programme for research, is taking the opposite approach. It wants to set up 'networks of excellence', which must involve separate entities in three different countries, two of which must be EC member or member-associated nations.

The open question is, which approach will attract the best researchers and promote the most fruitful interactions? Some researchers I met in Germany and France agree that the EC aim is noble. But they question the means to achieve it. If collaborations become too cumbersome — with the bureaucracy needed to cement them becoming an obstacle to good science and career development — the term may well gain too many negative connotations.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Biodefence gets a booster jab p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

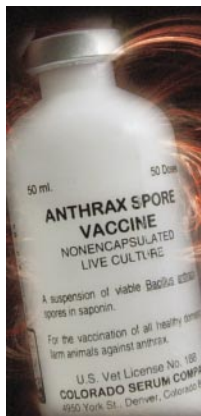
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Funding increases for biodefence research in the United States will have a sizable impact on the landscape for jobs, says Eugene Russo.

John Young's colleagues used to wonder why he worked on anthrax. "It was an intellectual curiosity, really, that drove the project," says Young, a professor of cancer research at the University of Wisconsin, Madison. Since the events of last autumn, Young's intellectual pursuits have acquired national security implications.

In 1999, Young, who originally trained as a retrovirologist, began investigating anthrax toxin receptors. He was attracted to the project in part because anthrax toxin and the avian retrovirus he had been studying use similar mechanisms for cell entry.

Driven by the joint interests of graduate student Ken Bradley in the Young lab and of postdoc Jeremy Mogridge in John Collier's lab at Harvard Medical School, the high-risk project went forward with limited resources. But it would pay big dividends. Last November, the group reported the first cloning of an anthrax toxin receptor (K. A. Bradley *et al.* *Nature* **414**, 225–229; 2001). The team submitted the paper to *Nature* just a few days before the first anthrax cases were reported in the United States. Since then, Young has devoted more time to anthrax and now collaborates with scientists from several fields.

Young's story may not be typical, but it exemplifies one way in which biodefence concerns have already affected research. The impact will intensify with the massive increase planned in federal US funding.

Support for basic research in biodefence through the National Institute of Allergy and Infectious Diseases (NIAID) is slated to increase by \$1.5 billion in fiscal year 2003, in addition to the \$270 million already budgeted. The boost is part of a proposed \$4.5-billion funding increase for general biodefence spending in 2003, up 319% from 2002.

The NIAID's funding boost is likely to affect the research landscape for microbiologists, cell

biologists, virologists and chemists. Indeed, some of these specialists have even begun to band together into 'biodefence interest groups' to share expertise and jointly apply for funding. Anthony Fauci, the NIAID's director, who has heard from a few such groups, is enthusiastic about the prospect of funding them because, he says, their multidisciplinary increases their chances of success.

BROAD-BASED STRATEGY

The anthrax attacks of last autumn, although tragic, had a relatively minor public-health impact. How can the planned huge funding boost be justified, especially if further attacks do not materialize in the near future?

There is one obvious answer: the need to be prepared. The other, according to Fauci, is the need not only to bolster biodefence but to improve scientists' understanding of both the immune system and the pathogenesis of infectious diseases in general. "We want to include bioterrorism in the big umbrella of emerging and re-emerging infections," says Fauci.

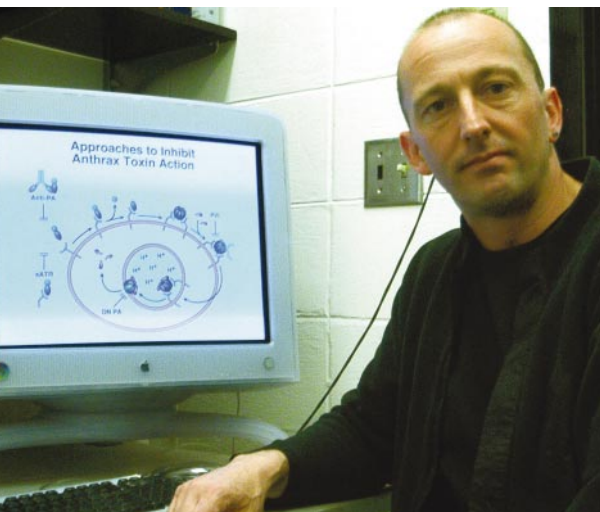
Research targets will go well beyond the designated 'category A' bioterrorist agents such as anthrax, botulism, plague and smallpox. Although these have been the focus of media attention, Fauci emphasizes the importance of bolstering manpower and infrastructure, and fostering general study of host-defence interactions, immunotherapy and adaptive immunity. Work on diagnostics, therapeutics and vaccines is also likely to have broader applications.

"The idea is to design a platform technology for application against any pathogen," says Adel Mahmoud, head of vaccine research at drug firm Merck. The company is collaborating with the NIAID on the development of an HIV vaccine.

The biodefence research agenda, in fact, has several similarities with the research onslaught that resulted from the HIV epidemic in the 1980s. Fauci believes that that mammoth task will provide lessons for this one. As with the HIV/AIDS research strategy, he foresees extensive collaborations with industrial partners, partly solicited through challenge grants for bioterrorism-related research projects.

Over the next few years the NIAID will also establish 6–10 centres of excellence in emerging diseases and biodefence at universities and institutes. The first of these will be funded from the 2003 budget.

John Young: curiosity about anthrax proved to be timely.





Genomics still holds the key

Already a hot area, genomics is to be a key component of the National Institutes of Health's anti-bioterrorism basic-research strategy. Although all the 'category A' agents, such as anthrax and smallpox, have been sequenced or targeted for sequencing, finding genetic differences between strains will be essential, says Claire Fraser, director of The Institute for Genomic Research (TIGR) in Rockville, Maryland.

"Having a single genome sequence for an important pathogen is probably not enough if you want to understand all of the variability you're going to find out there in nature," she says.

Last November, TIGR received \$25 million from the National Institute of Allergy and Infectious Diseases (NIAID) to establish the Pathogen Functional Genomics Resource Center. Its aim is to determine the biological function of pathogen genes and their protein products. TIGR and its collaborators recently submitted a proposal to NIAID to work on several anthrax strains.

TIGR has not yet hired more researchers, but bioinformatics experts are constantly in demand for all sequencing-related projects. "Bioinformatics is one area where we've never been fully staffed," says Fraser. **E.R.**

On the hunt: the US is ploughing millions of dollars into the search for countermeasures to biological warfare.

AP

The centres will have biosafety level 3 and level 4 handling facilities to contain the most dangerous pathogens, facilities that are currently in short supply. As with many of the anti-bioterrorism grant proposals, the NIAID will use an accelerated form of peer review to decide which centres to fund.

As with the HIV/AIDS research programme, the NIAID's funding strategy will ideally be broad enough from the start to attract established immunologists and infectious-disease specialists. "It's going to be a combination of programmatic direction, together with *de novo*, investigator-initiated projects," says Fauci.

Dennis Kasper, a Harvard Medical School professor of microbiology and molecular genetics, suggests that attracting top people to HIV research took some time. He hopes that this time more top scientists will get involved sooner. "You've got to have a way to attract the top-level scientists into the field," he says.

INFECTIOUS ENTHUSIASM

Fauci hopes not only to encourage infectious-disease investigators to make their work more relevant to biodefence, but also to help to train young students and postdocs with a "substantial recruitment and training effort". Early signs indicate that students may have already started to respond.

Applications to graduate biology programmes at major universities have increased 15% this year according to Kasper, an executive dean for academic programmes at Harvard Medical School. Kasper attributes the increase to a sagging computer-sciences job market and a general renewed interest in biology since bioterrorism hit the news. "Whether that will translate into more graduate students in microbiology, I hope so, but I don't know," he says.

Senior scientists are already showing interest in biodefence-related basic research. Young recently visited oncovirologist Peter Vogt at the Scripps Research Institute in San Diego, California. To Young's surprise, Vogt announced that he had started working on anthrax toxins. "That's all he wanted to talk about," recalls Young.

Since late last year, Vogt has been participating in a biodefence interest group studying several bioterrorism agents. For years, Vogt has collaborated with Scripps chemists to develop high-throughput tests for specific cancer targets. It was only a small step from the search for small molecules that inhibit

biologically active compounds to the search for molecules that might inhibit bacterial toxins.

Last year, Vogt's collaborator Barry Sharpless, Scripps chemist and Nobel laureate, found a very potent enzyme inhibitor in the context of their cancer research. But they recognized its potential for more widespread use. "We started thinking about toxins before 11 September," says Vogt. But after that date, "the obvious candidate immediately became anthrax".

Vogt has become fascinated by specific actions of the toxin and is surprised by the small number of experts studying these actions. He contends that this is much more than derivative applied work. "These are fantastically interesting scientific questions that are really challenging," he says. ■

Eugene Russo is freelance science writer in Takoma Park, Maryland.

NIAID counter-bioterrorism research agenda

♦ www.niaid.nih.gov/dmid/pdf/biotresearchagenda.pdf

NIAID funding

♦ www.niaid.nih.gov/dmid/bioterrorism

Anthony Fauci sees a multidisciplinary approach as the key to success.

